GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 2765.41 Seconds
(without alignments)
14608.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....acctgattctggtcttcttt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1001	100.0	110000	6	AX093471	AX093471 Sequence
2		1001	100.0	110000	9	HSA246003	AJ246003 Homo sapi
3		1001	100.0	162692	9	CNS01DS8	AL121658 BAC seque
4	c	1001	100.0	185281	2	AC011232	AC011232 Homo sapi
5		403	40.3	3263	6	AX093472	AX093472 Sequence
6		403	40.3	3263	9	HSA246001	AJ246001 Homo sapi
7		403	40.3	5120	9	AB029006	AB029006 Homo sapi
8	c	200.6	20.0	172816	9	AC093899	AC093899 Homo sapi
9		199.8	20.0	195432	9	AC130403	AC130403 Homo sapi
10	c	199.8	20.0	209354	2	AC068752	AC068752 Homo sapi
11		197.8	19.8	163027	9	AC002549	AC002549 Homo sapi
12	c	197.8	19.8	163662	2	AC096508	AC096508 Homo sapi
13	c	197.8	19.8	191866	2	AC068676	AC068676 Homo sapi
14		196.8	19.7	119922	9	AC099804	AC099804 Homo sapi
15	c	195.8	19.6	46775	9	HSJ738A13	AL109801 Human DNA
16		195.6	19.5	182799	2	AC073526	AC073526 Homo sapi
17		195.4	19.5	130569	9	CNS08CBB	AL845552 Human chr
18	c	195.4	19.5	137506	9	AC004824	AC004824 Homo sapi
19	c	194.6	19.4	177300	9	AC096741	AC096741 Homo sapi
20		194.4	19.4	45373	9	AL158092	AL158092 Human DNA
21	c	194.4	19.4	173520	9	AL732423	AL732423 Human DNA
22	c	193.6	19.3	176498	9	AC003035	AC003035 Homo sapi
23		193.6	19.3	178908	9	AL157392	AL157392 Human DNA
24	c	193.6	19.3	189485	2	AC141268	AC141268 Homo sapi
25		193.4	19.3	81542	9	AC100830	AC100830 Homo sapi
26		193.4	19.3	111744	2	AC099686	AC099686 Homo sapi
27	c	193.4	19.3	150147	9	AC002543	AC002543 Homo sapi
28	c	193.4	19.3	202920	2	AC139392	AC139392 Rattus no
29		193.4	19.3	215829	2	AC131423	AC131423 Rattus no
30	c	193.2	19.3	82517	9	HS341D10	Z97985 Human DNA s
31	c	193	19.3	169476	9	CNS01DW7	AL136419 Human chr
32	c	192.8	19.3	140363	9	AC142331	AC142331 Pan trogl
33	c	192.8	19.3	172533	9	AC002381	AC002381 Human BAC
34		192.4	19.2	150489	2	AC069491	AC069491 Homo sapi
35		192.4	19.2	157574	9	AL592293	AL592293 Human DNA
36	c	192.4	19.2	160169	2	AC051664	AC051664 Homo sapi
37		192.4	19.2	163915	2	AC087451	AC087451 Homo sapi
38	c	192.4	19.2	156867	9	AP003733	AP003733 Homo sapi
39		192.4	19.2	178770	2	AC090725	AC090725 Homo sapi
40	c	192.4	19.2	181140	9	AP002793	AP002793 Homo sapi
41	c	192.4	19.2	205150	9	AC005274	AC005274 Homo sapi
42	c	192.2	19.2	145423	9	AP003086	AP003086 Homo sapi
43	c	192.2	19.2	161096	9	AC016257	AC016257 Homo sapi
44		192.2	19.2	167941	2	AC058804	AC058804 Homo sapi
45		192.2	19.2	185737	9	AC005488	AC005488 Homo sapi

ALIGNMENTS

RESULT 1
AX093471
LOCUS AX093471 110000 bp DNA linear PAT 30-MAR 2001
DEFINITION Sequence 1 from Patent WO0118198.
ACCESSION AX093471
VERSION AX093471.1 GI:13509911
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Weissenbach,J. and Hazan,J.
TITLE Cloning, expression and characterisation of the spg4 gene
responsible for the most frequent form of autosomal spastic

```
            paraplegia
  JOURNAL   Patent: WO 0118198-A 1 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     exon            9932. .10471
     intron          10472. .33718
     exon            33719. .33805
     intron          33806. .35748
     exon            35749. .35832
     intron          35833. .45022
     exon            45023. .45118
     intron          45119. .60863
     exon            60864. .61051
     intron          61052. .61927
     exon            61928. .62061
     intron          62062. .62344
     exon            62345. .62438
     intron          62439. .73173
     exon            73174. .73248
     intron          73249. .74633
     exon            74634. .74705
     intron          74706. .82788
     exon            82789. .82864
     intron          82865. .83102
     exon            83103. .83194
     intron          83195. .83334
     exon            83335. .83414
     intron          83415. .88129
     exon            88130. .88172
     intron          88173. .89561
     exon            89562. .89641
     intron          89642. .91162
     exon            91163. .91233
     intron          91234. .93443
     exon            93444. .93484
     intron          93485. .100599
     exon            100600. .102009
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match             100.0%;  Score 1001;  DB 6;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-181;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100000 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 100059

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100060 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 100119

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100120 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 100179

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100180 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 100239

Qy        241 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100240 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 100299

Qy        301 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100300 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 100359

Qy        361 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100360 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 100419

Qy        421 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100420 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 100479

Qy        481 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100480 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 100539

Qy        541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100540 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 100599

Qy        601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100600 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 100659

Qy        661 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100660 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 100719

Qy        721 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100720 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 100779

Qy        781 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100780 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 100839

Qy        841 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100840 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 100899

Qy        901 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100900 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 100959

Qy        961 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
              ||||||||||||||||||||||||||||||||||||||||||
Db     100960 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 101000

RESULT 2
HSA246003
LOCUS       HSA246003            110000 bp    DNA     linear   PRI 08-MAR-2000
DEFINITION  Homo sapiens Spast gene for spastin protein.
ACCESSION   AJ246003
VERSION     AJ246003.1  GI:6273492
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
```

```
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 110000)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The
            sequence is the result of the assembly of 2 BAC clones: R-336p14
            and
            563N4, respectively from RPCI-11 and CITB_978_SKB library.
FEATURES             Location/Qualifiers
     source          1..110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
                     /clone="R-336P14"
                     /clone_lib="RPCI-11"
                     /note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
                     , BAC 563N4 contains Spast gene: exon 4 to exon 17"
     gene            9932..102009
                     /gene="Spast"
     mRNA            9932..102009
                     /gene="Spast"
                     /product="spastin protein"
     exon            9932..10471
                     /gene="Spast"
                     /number=1
     5'UTR           9932..10056
                     /gene="Spast"
     CDS             join(10057..10471,33719..33805,35749..35832,45023..45118,
                     60864..61051,61928..62061,62345..62438,73174..73248,
                     74634..74705,82789..82864,83103..83194,83335..83414,
                     88130..88172,89562..89641,91163..91233,93444..93484,
                     100600..100722)
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60208.1"
                     /db_xref="GI:6273493"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     intron          10472..33718
                     /gene="Spast"
                     /number=1
     exon            33719..33805
                     /gene="Spast"
                     /number=2
     intron          33806..35748
                     /gene="Spast"
                     /number=2
     exon            35749..35832
                     /gene="Spast"
                     /number=3
     intron          35833..45022
                     /gene="Spast"
                     /number=3
     exon            45023..45118
                     /gene="Spast"
                     /number=4
     intron          45119..60863
                     /gene="Spast"
                     /number=4
     exon            60864..61051
                     /gene="Spast"
                     /number=5
     intron          61052..61927
                     /gene="Spast"
                     /number=5
     exon            61928..62061
                     /gene="Spast"
                     /number=6
     intron          62062..62344
                     /gene="Spast"
                     /number=6
     exon            62345..62438
                     /gene="Spast"
                     /number=7
     intron          62439..73173
                     /gene="Spast"
                     /number=7
     exon            73174..73248
                     /gene="Spast"
                     /number=8
     intron          73249..74633
                     /gene="Spast"
                     /number=8
     exon            74634..74705
                     /gene="Spast"
                     /number=9
     intron          74706..82788
                     /gene="Spast"
                     /number=9
     exon            82789..82864
                     /gene="Spast"
                     /number=10
     intron          82865..83102
                     /gene="Spast"
                     /number=10
     exon            83103..83194
                     /gene="Spast"
                     /number=11
     intron          83195..83334
                     /gene="Spast"
                     /number=11
     exon            83335..83414
                     /gene="Spast"
                     /number=12
     intron          83415..88129
                     /gene="Spast"
                     /number=12
     exon            88130..88172
                     /gene="Spast"
                     /number=13
     intron          88173..89561
                     /gene="Spast"
                     /number=13
     exon            89562..89641
                     /gene="Spast"
                     /number=14
     intron          89642..91162
                     /gene="Spast"
                     /number=14
     exon            91163..91233
                     /gene="Spast"
                     /number=15
     intron          91234..93443
                     /gene="Spast"
                     /number=15
     exon            93444..93484
                     /gene="Spast"
                     /number=16
     intron          93485..100599
```

```
                    /gene="Spast"
                    /number=16
     exon           100600..102009
                    /gene="Spast"
                    /number=17
     3'UTR          100721..102009
                    /gene="Spast"
     polyA_signal   101973..101978
                    /gene="Spast"
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match             100.0%;  Score 1001;  DB 9;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 9.7e-181;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
Db     100000 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 100059

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
Db     100060 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 100119

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
Db     100120 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 100179

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
Db     100180 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 100239

Qy        241 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 300
Db     100240 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 100299

Qy        301 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 360
Db     100300 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 100359

Qy        361 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 420
Db     100360 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 100419

Qy        421 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 480
Db     100420 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 100479

Qy        481 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
Db     100480 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 100539

Qy        541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
Db     100540 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 100599

Qy        601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 660
Db     100600 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 100659

Qy        661 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 720
Db     100660 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 100719

Qy        721 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 780
Db     100720 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 100779

Qy        781 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 840
Db     100780 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 100839

Qy        841 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 900
Db     100840 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 100899

Qy        901 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 960
Db     100900 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 100959

Qy        961 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db     100960 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 101000


RESULT 3
CNS01DS8
LOCUS       CNS01DS8               162692 bp    DNA     linear   PRI 18-APR-2002
DEFINITION  BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04
            of library RPCI-11 from chromosome 2 of Homo sapiens (Human).
ACCESSION   AL121658
VERSION     AL121658.4  GI:20218783
KEYWORDS    SPG4 genomic DNA interval.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162692)
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a novel AAA protein, is altered in the most frequent form
            of autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. (1999) In press
REFERENCE   2  (bases 1 to 162692)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Apr 19, 2002 this sequence version replaced gi:20160242.
FEATURES             Location/Qualifiers
     source          1..162692
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="563N04"
                     /clone_lib="RPCI-11"
BASE COUNT    44981 a  32017 c  32887 g  52792 t     15 others
ORIGIN

  Query Match             100.0%;  Score 1001;  DB 9;  Length 162692;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-181;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
Db      81260 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 81319

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
Db      81320 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 81379

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
Db      81380 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 81439

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
Db      81440 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 81499

Qy        241 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 300
```

```
Db      81500 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 81559

Qy        301 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81560 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 81619

Qy        361 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81620 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 81679

Qy        421 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81680 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 81739

Qy        481 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81740 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 81799

Qy        541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81800 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 81859

Qy        601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81860 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 81919

Qy        661 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81920 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 81979

Qy        721 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81980 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 82039

Qy        781 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82040 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 82099

Qy        841 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82100 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 82159

Qy        901 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      82160 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 82219

Qy        961 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
              |||||||||||||||||||||||||||||||||||||||||
Db      82220 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 82260


RESULT 4
AC011232/c
LOCUS       AC011232              185281 bp    DNA     linear   HTG 10-MAR-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC011232
VERSION     AC011232.7  GI:13270720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 10, 2001 this sequence version replaced gi:9799811.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0078E13
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 57%
            Sequencing vector: plasmid; 40%
            Chemistry: Dye-primer ET; 48% of reads
            Chemistry: Dye-terminator Big Dye; 50% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182798 bases at least Q40
            Consensus quality: 183538 bases at least Q30
            Consensus quality: 184045 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 184681; sum-of-contigs
            Quality coverage: 7.07 in Q20 bases; agarose-fp
            Quality coverage: 7.33 in Q20 bases; sum-of-contigs
            ---------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1985: contig of 1985 bp in length
            *     1986     2085: gap of unknown length
            *     2086     8183: contig of 6098 bp in length
            *     8184     8283: gap of unknown length
            *     8284    22741: contig of 14458 bp in length
            *    22742    22841: gap of unknown length
            *    22842    41446: contig of 18605 bp in length
            *    41447    41546: gap of unknown length
            *    41547    69182: contig of 27636 bp in length
            *    69183    69282: gap of unknown length
            *    69283   112131: contig of 42849 bp in length
            *   112132   112231: gap of unknown length
            *   112232   185281: contig of 73050 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185281
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-78E13"
     misc_feature    1. .1985
                     /note="assembly_name:Contig11"
     misc_feature    2086. .8183
                     /note="assembly_name:Contig12"
     misc_feature    8284. .22741
                     /note="assembly_name:Contig13"
     misc_feature    22842. .41446
                     /note="assembly_name:Contig14"
     misc_feature    41547. .69182
                     /note="assembly_name:Contig15"
     misc_feature    69283. .112131
                     /note="assembly_name:Contig16"
     misc_feature    112232. .185281
                     /note="assembly_name:Contig17"
BASE COUNT    52719 a  36957 c  37187 g  57816 t    602 others
ORIGIN

  Query Match           100.0%;  Score 1001;  DB 2;  Length 185281;
  Best Local Similarity  100.0%;  Pred. No. 9.3e-181;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      64473 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 64414

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64413 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 64354

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64353 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 64294

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64293 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 64234

Qy        241 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 300
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64233 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 64174

Qy        301 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 360
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64173 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 64114

Qy        361 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 420
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64113 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 64054

Qy        421 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 480
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      64053 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 63994

Qy        481 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63993 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 63934

Qy        541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63933 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 63874

Qy        601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAAATAAAACGCAGCGTC 660
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63873 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAAATAAAACGCAGCGTC 63814

Qy        661 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 720
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63813 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 63754

Qy        721 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 780
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63753 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 63694

Qy        781 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 840
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63693 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 63634

Qy        841 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 900
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63633 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 63574

Qy        901 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 960
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      63573 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 63514

Qy        961 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
               |||||||||||||||||||||||||||||||||||||||||
Db      63513 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 63473


RESULT 5
AX093472
LOCUS       AX093472               3263 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 2 from Patent WO0118198.
ACCESSION   AX093472
VERSION     AX093472.1  GI:13509912
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 2 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS); (FR)
FEATURES             Location/Qualifiers
     source          1..3263
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      956 a    664 c    703 g    940 t
ORIGIN

  Query Match           40.3%;  Score 403;  DB 6;  Length 3263;
  Best Local Similarity  100.0%;  Pred. No. 5.7e-67;
  Matches  403;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1852 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 1911

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1912 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 1971

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1972 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 2031

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2032 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 2091

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2092 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 2151

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2152 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 2211

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
               |||||||||||||||||||||||||||||||||||||||||||
Db       2212 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 2254


RESULT 6
HSA246001
LOCUS       HSA246001              3263 bp    mRNA    linear   PRI 10-MAR-2000
DEFINITION  Homo sapiens mRNA for spastin protein (Spast gene).
ACCESSION   AJ246001
VERSION     AJ246001.1  GI:6273490
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
```

```
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Doufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 3263)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p22"
     gene            1. .3263
                     /gene="Spast"
     CDS             126. .1976
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60141.1"
                     /db_xref="GI:6273491"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYPSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     polyA_signal    3227. .3232
                     /gene="Spast"
BASE COUNT      956 a    664 c    703 g    940 t
ORIGIN

  Query Match             40.3%;  Score 403;  DB 9;  Length 3263;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-67;
  Matches  403;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
              ||||||||| ||||| |||||| |||||.|||||||||||| |||||.||||||||||||
Db       1852 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 1911

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
              |||||.|.|||||||.|||| ||||||||||||||||||||||||.||| .||| ||||
Db       1912 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 1971

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
              ||| |||||||||||||||||||||||||||||||||||.|||||||||||||||||||
Db       1972 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 2031

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
              ||||||||||||||||||||||||||||||||||||||||||||| |||| |||||||| ||
Db       2032 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 2091

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2092 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 2151

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
              |||||||||||| ||| |||||| |||||.|||||||||||| |||||.||||||||||||
Db       2152 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 2211

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
              |||||.|||||||||||.|||| |||||||||||||||||||
Db       2212 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 2254


RESULT 7
AB029006
LOCUS       AB029006               5120 bp    mRNA    linear   PRI 04-AUG-1999
DEFINITION  Homo sapiens mRNA for KIAA1083 protein, complete cds.
ACCESSION   AB029006
VERSION     AB029006.1  GI:5689502
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kikuno,R., Nagase,T., Ishikawa,K., Hirosawa,M., Miyajima,N.,
            Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XIV. The complete sequences of 100 new cDNA clones from brain which
            code for large proteins in vitro
  JOURNAL   DNA Res. 6 (3), 197-205 (1999)
  MEDLINE   99397452
   PUBMED   10470851
REFERENCE   2  (bases 1 to 5120)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1. .5120
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="hj07423"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
     gene            1. .5120
                     /gene="KIAA1083"
     CDS             222. .1976
                     /gene="KIAA1083"
                     /codon_start=1
                     /product="KIAA1083 protein"
                     /protein_id="BAA83035.1"
                     /db_xref="GI:5689503"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYPSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLESGAVPKRKDPLTHTSNSLPRSK
                     TVMKTGSAGLSGHHRAPSYSGLSMVSGVKQGSGPAPTTHKGTPKTNRTNKPSTPTTAT
                     RKKKDLKNFRNVDSNLANLIMNEIVDNGTAVKFDDIAGQDLAKQALQEIVILPSLRPE
                     LFTGLRAPARGLLLFGPPGNGKTMLAKAVAAESNATFFNISAASLTSKYVGEGEKLVR
                     ALFAVARELQPSIIFIDEVDSLLCERREGEHDASRRLKTEFLIEFDGVQSAGDDRVLV
                     MGATNRPQELDEAVLRRFIKRVYVSLPNEETRLLLLKNLLCKQGSPLTQKELAQLARM
                     TDGYSGSDLTALAKDAALGPIRELKPEQVKNMSASEMRNIRLSDFTESLKKIKRSVSP
                     QTLEAYIRWNKDFGDTTV"
BASE COUNT     1532 a    943 c   1077 g   1568 t
ORIGIN

  Query Match             40.3%;  Score 403;  DB 9;  Length 5120;
  Best Local Similarity   100.0%;  Pred. No. 5.5e-67;
  Matches  403;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
```

```
Db       1852 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 1911

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718

Db       1912 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 1971

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778

Db       1972 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 2031

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838

Db       2032 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 2091

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898

Db       2092 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 2151

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958

Db       2152 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 2211

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001

Db       2212 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 2254


RESULT 8
AC093899/c
LOCUS       AC093899               172816 bp    DNA     linear   PRI 12-JUN-2002
DEFINITION  Homo sapiens BAC clone RP11-724O16 from 2, complete sequence.
ACCESSION   AC093899 AC068884
VERSION     AC093899.3  GI:18497265
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172816)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
    PUBMED  9847074
REFERENCE   2  (bases 1 to 172816)
  AUTHORS   Pearman,C., Haakenson,W. and Boyer,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-724O16
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 172816)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 172816)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-FEB-2002) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 172816)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Feb 5, 2002 this sequence version replaced gi:15625013.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0724O16
                Drafting Center: WIBR
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO. For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.C. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            Actual start of this clone is at base position 1 of RP11-724O16;
            actual end is at base position 172816 of RP11-724O16.

            Unresolved simple sequence repeat from base position 22700 to
            25900.

            The sequence of AC068884 has been incorporated into AC093899.
FEATURES             Location/Qualifiers
     source          1. .172816
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-724O16"
                     /clone_lib="RPCI-11"
     repeat_region   19. .1140
                     /rpt_family="L1"
     repeat_region   1141. .1437
                     /rpt_family="Alu"
     repeat_region   1438. .2657
                     /rpt_family="L1"
     repeat_region   2658. .2729
                     /rpt_family="L1"
     repeat_region   2730. .3320
                     /rpt_family="L1"
     repeat_region   3321. .3642
                     /rpt_family="Alu"
     repeat_region   3643. .4184
                     /rpt_family="L1"
     repeat_region   4185. .4996
                     /rpt_family="ERV1"
     repeat_region   5011. .5263
                     /rpt_family="L1"
     repeat_region   5285. .5907
                     /rpt_family="CR1"
     repeat_region   6895. .7078
                     /rpt_family="L1"
     repeat_region   7103. .7421
```

```
                 /rpt_family="Alu"
repeat_region    7807..7952
                 /rpt_family="L1"
repeat_region    7981..8289
                 /rpt_family="Alu"
repeat_region    8360..8551
                 /rpt_family="L1"
repeat_region    8622..8718
                 /rpt_family="MIR"
repeat_region    8794..8825
                 /rpt_family="(TTTTC)n"
repeat_region    8840..9134
                 /rpt_family="Alu"
repeat_region    9136..9174
                 /rpt_family="Alu"
repeat_region    9861..9881
                 /rpt_family="(A)n"
repeat_region    9882..10166
                 /rpt_family="Alu"
repeat_region    16502..16556
                 /rpt_family="L2"
repeat_region    16886..17165
                 /rpt_family="L1"
repeat_region    17179..17482
                 /rpt_family="Alu"
repeat_region    18025..18119
                 /rpt_family="MIR"
repeat_region    18177..18471
                 /rpt_family="Alu"
repeat_region    19312..19441
                 /rpt_family="MER1_type"
repeat_region    19731..20040
                 /rpt_family="MaLR"
repeat_region    20403..20472
                 /rpt_family="L2"
repeat_region    21286..21344
                 /rpt_family="(CA)n"
repeat_region    22724..23977
                 /rpt_family="(TA)n"
repeat_region    24005..24171
                 /rpt_family="(TA)n"
repeat_region    24196..25007
                 /rpt_family="(TA)n"
repeat_region    25030..25205
                 /rpt_family="(TA)n"
repeat_region    25223..25398
                 /rpt_family="(TA)n"
repeat_region    25399..25455
                 /rpt_family="AT_rich"
repeat_region    25456..25919
                 /rpt_family="(TA)n"
repeat_region    25920..26206
                 /rpt_family="Alu"
repeat_region    26893..26917
                 /rpt_family="AT_rich"
repeat_region    27194..27497
                 /rpt_family="L1"
repeat_region    27498..27814
                 /rpt_family="MER2_type"
repeat_region    27815..28151
                 /rpt_family="L1"
repeat_region    28152..28278
                 /rpt_family="Alu"
repeat_region    28279..28485
                 /rpt_family="L1"
repeat_region    28795..28829
                 /rpt_family="AT_rich"
repeat_region    28986..29100
                 /rpt_family="L2"
repeat_region    29237..29349
                 /rpt_family="L1"
repeat_region    29350..29484
                 /rpt_family="Alu"
repeat_region    29485..29589
                 /rpt_family="L1"
repeat_region    29811..29874
                 /rpt_family="L2"
repeat_region    29875..30284
                 /rpt_family="L1"
repeat_region    30320..30522
                 /rpt_family="L1"
repeat_region    30523..30847
                 /rpt_family="Alu"
repeat_region    30848..30882

Query Match             20.0%;  Score 200.6;  DB 9;  Length 172616;
Best Local Similarity   60.8%;  Pred. No. 1.2e-28;
Matches  398;  Conservative    0;  Mismatches  249;  Indels    9;  Gaps    4;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64
Db      26223 TTATAATTATCATTATTGGCTGGGCGCAGTGGCTCATGTCTGTAATCCCAGCACTTTGGG 26164

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
Db      26163 AGGCTGAGGTGGGCAGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAATATAGT 26104

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184
Db      26103 GAAACTTCGTCTCTACTAAAAATACAAAAATTAGCCAAGCATGGTGGTGTGTGCCTGTAA 26044

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
Db      26043 TCCCAGCTACTGGGGAGTCTGAGGCAGGAGAACTGCTTGAACCTGAGAGGTAGAGGTTGC 25984

Qy        245 ATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACTTTTC 304
Db      25983 AT---TGAGTGGAGAGAGTGCCACTGCCCTCCAGTCTGGGTAACAGAGCAAGACTCCATC 25927

Qy        305 TATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTAATAC 364
Db      25926 TAAAAAATATATATATATATAAAATATATAAATATATATAAATATAAATATATAATATAA 25867

Qy        365 CTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCTTGTG 424
Db      25866 ATATATAATATAATATATATAATATATTATATAATATATAATATATATAATATACATATA 25807

Qy        425 TAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTA---AGAATACA 481
Db      25806 TA-AATATATATATAAAATATATAAATAAAAAATAAATATAAATATAAATATATATAACA 25748

Qy        482 TACACGTATATTTTTATAACAT-TAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
Db      25747 TATAAATATATATAAATAAATATATAATATATAAAATATAAATATATATAATATATATAA 25688

Qy        541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
Db      25687 AATTTATATATATTATATATATTATATATATAAATATATATATAAAATTTATAAATATAT 25628

Qy        601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCA 655
Db      25627 ATACAATATATAAAATTTATATATATATTTACATATATAAAATATATAAAATGTA 25573


RESULT 9
AC130403
LOCUS       AC130403            195432 bp    DNA     linear   PRI 27-DEC-2002
DEFINITION  Homo sapiens 12 BAC RP11-674N14 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC130403
VERSION     AC130403.3  GI:24431574
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

REFERENCE 1 (bases 1 to 195432)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbaria,J., Benton,J., Birage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S., Escotto,M., Falis,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,C.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,C., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,C., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,N., Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C., Shoosh tari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,R., Stone,R., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 195432)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 195432)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 195432)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 195432)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-DEC-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 31, 2002 this sequence version replaced gi:23603880.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

FEATURES	Location/Qualifiers
source	1..195432
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="12"
	/clone="RP11-674N14"
misc_feature	1..2004
	/note="overlaps bases 11278..13281 of clone AC137929"
	/function="clone overlap"
repeat_region	complement(419..927)
	/rpt_family="LTR40a"
repeat_region	928..953
	/rpt_family="(CA)n"
repeat_region	1026..1072
	/rpt_family="(TTATA)n"
repeat_region	complement(1802..1929)
	/rpt_family="MLT1J"
repeat_region	2381..2592
	/rpt_family="L2"
repeat_region	2598..2822
	/rpt_family="AluSg/x"
repeat_region	2892..3614
	/rpt_family="L2"
repeat_region	3615..4032
	/rpt_family="MSTA"
repeat_region	4033..4087
	/rpt_family="L2"
repeat_region	complement(4094..4456)
	/rpt_family="MER66A"
repeat_region	4492..4963
	/rpt_family="L2"
repeat_region	complement(4966..5269)
	/rpt_family="AluY"
repeat_region	5607..5741
	/rpt_family="MIR"
repeat_region	complement(5778..5974)
	/rpt_family="MER3"
repeat_region	complement(6110..6476)

```
                  /rpt_family="L1PB1"
     repeat_region  6477..6510
                  /rpt_family="(TG)n"
     repeat_region  6510..6688
                  /rpt_family="(TA)n"
     repeat_region  complement(6511..6702)
                  /rpt_family="L1PB1"
     repeat_region  6703..6762
                  /rpt_family="(TA)n"
     repeat_region  complement(6775..7080)
                  /rpt_family="L1PB1"
     repeat_region  7063..8073
                  /rpt_family="L1P5"
     repeat_region  8079..8224
                  /rpt_family="L1P5"
     repeat_region  complement(8344..8577)
                  /rpt_family="MLT1K"
     repeat_region  8666..8708
                  /rpt_family="(TATG)n"
     repeat_region  8779..8877
                  /rpt_family="L2"
     repeat_region  complement(9302..9463)
                  /rpt_family="FRAM"
     repeat_region  complement(9468..9507)
                  /rpt_family="MER116"
     repeat_region  9558..9839
                  /rpt_family="AluSp"
     repeat_region  9840..9883
                  /rpt_family="(TAAA)n"
     repeat_region  complement(10135..10518)
                  /rpt_family="MLT1B"
     repeat_region  10795..10822
                  /rpt_family="AT_rich"
     repeat_region  11068..11241
                  /rpt_family="L1P5"
     repeat_region  11272..11509
                  /rpt_family="L1PA12"
     repeat_region  11641..11789
                  /rpt_family="L1P3"
     repeat_region  12052..12332
                  /rpt_family="AluSg"
     repeat_region  12333..12383
                  /rpt_family="(TAAA)n"
     repeat_region  14057..14943

  Query Match             20.0%;  Score 199.8;  DB 9;  Length 195432;
  Best Local Similarity   88.9%;  Pred. No. 1.7e-28;
  Matches  216;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62
                 ||||  ||||  |  |||||||| ||||||||||  |||||  |  |||  |||   |  ||
Db      44699 TAGAAAACTCCATATTCTAGGCTGGGCAAGGTGGCTCACATCTGTAATCCCACTACTTTG 44758

Qy         63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122
              ||||| ||   ||| | ||||  ||||| |||  ||||| |||  ||  ||||  ||   ||
Db      44759 GGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATG 44818

Qy        123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182
               | |||||| |||||  ||||| |||||  ||||| ||||   |    |   ||| |||
Db      44819 GTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTCGCATGTGCCTGT 44878

Qy        183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242
              ||||||||||||||||||||||||| ||| |||||||||||||||   |||||    ||||
Db      44879 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTT 44938

Qy        243 GCA 245
              |||
Db      44939 GCA 44941

RESULT 10
AC068752/c
LOCUS       AC068752          209354 bp    DNA     linear   HTG 12-AUG-2000
DEFINITION  Homo sapiens chromosome 12 clone RP11-674N14, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC068752
VERSION     AC068752.2  GI:9798052
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 209354)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 209354)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Aug 12, 2000 this sequence version replaced gi:7717179.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ------------- Project Information -------------
            Center project name: H_NH0674N14
            ------------- Summary Statistics  -------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 189469 bases at least Q40
            Consensus quality: 194865 bases at least Q30
            Consensus quality: 197585 bases at least Q20
            Insert size: 205000; agarose-fp
            Insert size: 205754; sum-of-contigs
            Quality coverage: 3.54 in Q20 bases; agarose-fp
            Quality coverage: 3.63 in Q20 bases; sum-of-contigs
            -----------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *         1      1492: contig of 1492 bp in length
            *      1493      1592: gap of unknown length
            *      1593      3468: contig of 1876 bp in length
            *      3469      3568: gap of unknown length
            *      3569      6021: contig of 2453 bp in length
            *      6022      6121: gap of unknown length
            *      6122      8350: contig of 2229 bp in length
            *      8351      8450: gap of unknown length
            *      8451      9734: contig of 1284 bp in length
            *      9735      9834: gap of unknown length
            *      9835     12594: contig of 2760 bp in length
            *     12595     12694: gap of unknown length
            *     12695     14714: contig of 2020 bp in length
            *     14715     14814: gap of unknown length
            *     14815     17394: contig of 2580 bp in length
            *     17395     17494: gap of unknown length
            *     17495     20876: contig of 3382 bp in length
            *     20877     20976: gap of unknown length
            *     20977     24353: contig of 3377 bp in length
            *     24354     24453: gap of unknown length
            *     24454     27915: contig of 3462 bp in length
            *     27916     28015: gap of unknown length
            *     28016     31613: contig of 3598 bp in length
            *     31614     31713: gap of unknown length
```

```
        *    31714    34887: contig of 3174 bp in length
        *    34888    34987: gap of unknown length
        *    34988    38751: contig of 3764 bp in length
        *    38752    38851: gap of unknown length
        *    38852    42254: contig of 3403 bp in length
        *    42255    42354: gap of unknown length
        *    42355    46023: contig of 3669 bp in length
        *    46024    46123: gap of unknown length
        *    46124    50416: contig of 4293 bp in length
        *    50417    50516: gap of unknown length
        *    50517    54131: contig of 3615 bp in length
        *    54132    54231: gap of unknown length
        *    54232    58969: contig of 4738 bp in length
        *    58970    59069: gap of unknown length
        *    59070    64220: contig of 5151 bp in length
        *    64221    64320: gap of unknown length
        *    64321    68482: contig of 4162 bp in length
        *    68483    68582: gap of unknown length
        *    68583    74642: contig of 6060 bp in length
        *    74643    74742: gap of unknown length
        *    74743    80046: contig of 5304 bp in length
        *    80047    80146: gap of unknown length
        *    80147    85068: contig of 4922 bp in length
        *    85069    85168: gap of unknown length
        *    85169    91299: contig of 6131 bp in length
        *    91300    91399: gap of unknown length
        *    91400    96694: contig of 5295 bp in length
        *    96695    96794: gap of unknown length
        *    96795   102297: contig of 5503 bp in length
        *   102298   102397: gap of unknown length
        *   102398   110536: contig of 8139 bp in length
        *   110537   110636: gap of unknown length
        *   110637   116557: contig of 5921 bp in length
        *   116558   116657: gap of unknown length
        *   116658   124265: contig of 7608 bp in length
        *   124266   124365: gap of unknown length
        *   124366   133628: contig of 9263 bp in length
        *   133629   133728: gap of unknown length
        *   133729   141136: contig of 7408 bp in length
        *   141137   141236: gap of unknown length
        *   141237   153610: contig of 12374 bp in length
        *   153611   153710: gap of unknown length
        *   153711   165483: contig of 11773 bp in length
        *   165484   165583: gap of unknown length
        *   165584   176838: contig of 11255 bp in length
        *   176839   176938: gap of unknown length
        *   176939   192125: contig of 15187 bp in length
        *   192126   192225: gap of unknown length
        *   192226   209354: contig of 17129 bp in length.
FEATURES             Location/Qualifiers
     source          1..209354
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-674N14"
     misc_feature    1..1492
                     /note="assembly_name:Contig37"
     misc_feature    1593..3468
                     /note="assembly_name:Contig38"
     misc_feature    3569..6021
                     /note="assembly_name:Contig39"
     misc_feature    6122..8350
                     /note="assembly_name:Contig40"
     misc_feature    8451..9734
                     /note="assembly_name:Contig41"
     misc_feature    9835..12594
                     /note="assembly_name:Contig42"
     misc_feature    12695..14714
                     /note="assembly_name:Contig43"
     misc_feature    14815..17394
                     /note="assembly_name:Contig44
                     clone_end:SP6
                     vector_side:left"
     misc_feature    17495..20876
                     /note="assembly_name:Contig45"
     misc_feature    20977..24353
                     /note="assembly_name:Contig46"
     misc_feature    24454..27915
                     /note="assembly_name:Contig47"
     misc_feature    28016..31613
                     /note="assembly_name:Contig48"
     misc_feature    31714..34887
                     /note="assembly_name:Contig49"
     misc_feature    34988..38751
                     /note="assembly_name:Contig50"
     misc_feature    38852..42254
                     /note="assembly_name:Contig51"
     misc_feature    42355..46023
                     /note="assembly_name:Contig52"
     misc_feature    46124..50416
                     /note="assembly_name:Contig53"
     misc_feature    50517..54131
                     /note="assembly_name:Contig54"
     misc_feature    54232..58969
                     /note="assembly_name:Contig55"
     misc_feature    59070..64220
                     /note="assembly_name:Contig56"
     misc_feature    64321..68482
                     /note="assembly_name:Contig57"
     misc_feature    68583..74642
                     /note="assembly_name:Contig58"
     misc_feature    74743..80046
                     /note="assembly_name:Contig59"
     misc_feature    80147..85068
                     /note="assembly_name:Contig60"
     misc_feature    85169..91299
                     /note="assembly_name:Contig61"
     misc_feature    91400..96694
                     /note="assembly_name:Contig62"
     misc_feature    96795..102297
                     /note="assembly_name:Contig63"
     misc_feature    102398..110536
                     /note="assembly_name:Contig64"
     misc_feature    110637..116557
                     /note="assembly_name:Contig65"
     misc_feature    116658..124265
                     /note="assembly_name:Contig66"
     misc_feature    124366..133628
                     /note="assembly_name:Contig67
                     clone_end:T7
                     vector_side:left"
     misc_feature    133729..141136
                     /note="assembly_name:Contig68"
     misc_feature    141237..153610
                     /note="assembly_name:Contig69"

  Query Match             20.0%;  Score 199.8;  DB 2;  Length 209354;
  Best Local Similarity   88.9%;  Pred. No. 1.6e-26;
  Matches  216;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62
Db     202114 TAGAAAACTCCATATTCTAGGCTGGGCAAGGTGGCTCACATCTGTAATCCCACTACTTTG 202055

Qy         63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122
Db     202054 GGAGGCTGAGGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATG 201995

Qy        123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182
Db     201994 GTGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTCGCATGTGCCTGT 201935

Qy        183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242
Db     201934 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCTTGAACCCAGGAGGCGGAGGTT 201875
```

```
Qy        243 GCA 245
              ||
Db     201874 GCA 201872


RESULT 11
AC002549
LOCUS       AC002549              163027 bp    DNA     linear   PRI 27-MAR-1998
DEFINITION  Homo sapiens Xp22 BAC GS-377014 (Genome Systems Human BAC library)
            complete sequence.
ACCESSION   AC002549
VERSION     AC002549.1  GI:2992475
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163027)
  AUTHORS   Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,C.,
            Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
            Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,L.,
            Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
            Lu,C., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D.,
            Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,C.,
            Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163027)
  AUTHORS   Chiu,M.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-1997) Molecular and Human Genetics, Baylor
            College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163027)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-MAR-1998) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 27, 1998 this sequence version replaced gi:2360513.
            Sequencing is completed to a minimum standard of double strand
            coverage with a minimum of 2 clones and 2 reads with no ambiguities
            or 2 chemistries with a minimum of 2 clones and 3 reads with no
            ambiguities. If the sequence quality does not meet this standard,
            it will be indicated in the annotation.

            The repeat regions shown were identified using RepeatMasker by
            Adrian Smit.

            Sequence similarities were identified using Powerblast by Jinghui
            Zhang.

            Exon/Intron boundaries of identified genes were chosen if there
            were canonical splice junctions that maintained sequence continuity
            across the splice junctions.
FEATURES             Location/Qualifiers
     source          1..163027
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xp22"
                     /clone="GS-377014"
                     /clone_lib="Genome Systems Human BAC library"
     misc_feature    1..171
                     /note="overlap bases 123033-123203 in AC003036"
     repeat_region   525..823
                     /rpt_family="AluSx"
     repeat_region   891..1742
                     /rpt_family="L1M4"
     repeat_region   1776..1970
                     /rpt_family="L1PA13"
     repeat_region   2071..2159
                     /rpt_family="(TA)n"
     repeat_region   2232..2461
                     /rpt_family="AluJb"
     repeat_region   2480..2535
                     /rpt_family="POLY_A"
     repeat_region   complement(2541..2741)
                     /rpt_family="MER2"
     repeat_region   3707..3915
                     /rpt_family="MER58A"
     repeat_region   4261..4562
                     /rpt_family="AluSx"
     repeat_region   4704..4916
                     /rpt_family="MER20"
     repeat_region   complement(7185..7450)
                     /rpt_family="AluSg"
     repeat_region   7669..7708
                     /rpt_family="MIR"
     repeat_region   8605..8646
                     /rpt_family="AT_rich"
     repeat_region   complement(9401..9702)
                     /rpt_family="AluJb"
     repeat_region   9764..9950
                     /rpt_family="MER58B"
     repeat_region   10044..10306
                     /rpt_family="L1MB6"
     repeat_region   10309..10438
                     /rpt_family="AluSx"
     repeat_region   10440..10466
                     /rpt_family="(CAAA)n"
     repeat_region   10467..10651
                     /rpt_family="AluSx"
     repeat_region   10669..10941
                     /rpt_family="L1MB3"
     repeat_region   11888..12010
                     /rpt_family="L2"
     repeat_region   12405..12714
                     /rpt_family="L2"
     repeat_region   complement(14411..14713)
                     /rpt_family="AluSx"
     repeat_region   15004..15225
                     /rpt_family="MER20"
     repeat_region   17072..17147
                     /rpt_family="L2"
     repeat_region   18312..18457
                     /rpt_family="MIR"
     repeat_region   complement(18528..18821)
                     /rpt_family="AluJo"
     repeat_region   complement(19102..19403)
                     /rpt_family="AluSx"
     repeat_region   complement(19406..19587)
                     /rpt_family="MER33"
     repeat_region   complement(19666..19752)
                     /rpt_family="MER3"
     misc_feature    20048..20304
                     /note="Region: W04440 za43h03.r1 Soares fetal liver spleen
                     1NFLS Homo sapiens cDNA clone 295349 5'"
     repeat_region   complement(20265..20378)
                     /rpt_family="L1MA1"
     repeat_region   20439..20574
                     /rpt_family="MER33"
     repeat_region   complement(20578..20619)
                     /rpt_family="AT_rich"
     misc_feature    complement(21136..21502)
                     /note="Region: R98637 yq69h08.s1 Soares fetal liver spleen
                     1NFLS Homo sapiens cDNA clone 201087 3'"
     STS             21351..21502
                     /standard_name="WI-12677, Chr. -, Homo sapiens"
                     /db_xref="dbSTS:21718"
     repeat_region   23674..23770
                     /rpt_family="(CA)n"
     STS             23689..23816
                     /standard_name="AFMa067yd1, Chr. X, Homo sapiens"
                     /db_xref="dbSTS:29765"
```

```
     repeat_region   24776..25005
                     /rpt_family="L1MD3"
     repeat_region   25007..25295
                     /rpt_family="AluJo"
     repeat_region   25303..25344
                     /rpt_family="POLY_A"
     repeat_region   25918..26006
                     /rpt_family="L2"
     repeat_region   complement(26409..26625)
                     /rpt_family="AluJo"
     repeat_region   26983..27192
                     /rpt_family="MLT1F"
     repeat_region   29007..29429
                     /rpt_family="MLT1C"
     STS             29637..29722
                     /standard_name="DXS1313"
     repeat_region   30740..30771
                     /rpt_family="POLY_A"
     repeat_region   complement(31044..31336)
                     /rpt_family="AluSx"
     repeat_region   32605..32971
                     /rpt_family="MLT1A1"
     repeat_region   33700..33863
                     /rpt_family="MLT1C"
     repeat_region   33864..34165
                     /rpt_family="AluSx"
     repeat_region   34166..34197
                     /rpt_family="(CAAAA)n"
     repeat_region   34210..34481
                     /rpt_family="MLT1C"
     repeat_region   complement(34829..35177)
                     /rpt_family="MLT1B"
     repeat_region   36413..36519
                     /rpt_family="MER20"
     repeat_region   complement(36567..36706)
                     /rpt_family="MER20"
     repeat_region   complement(39685..39987)
                     /rpt_family="AluSg"
     repeat_region   40270..40647
                     /rpt_family="L1MB4"
     repeat_region   41624..41651
                     /rpt_family="AT_rich"
     repeat_region   complement(42747..43322)
                     /rpt_family="L1MD3"
     repeat_region   43727..43910
                     /rpt_family="MLT1C"
     repeat_region   44205..44499
                     /rpt_family="AluSx"
     repeat_region   44704..44780
                     /rpt_family="L1MB7"
     repeat_region   44782..45095
                     /rpt_family="MER2"
     repeat_region   45127..45474
                     /rpt_family="L1MB7"
     repeat_region   complement(46188..46488)
                     /rpt_family="AluJo"
     repeat_region   complement(46489..47116)
                     /rpt_family="L1MB2"
     repeat_region   complement(47166..47209)
                     /rpt_family="MIR"
     repeat_region   47211..47553
                     /rpt_family="MLT1C"

 Query Match             19.8%;  Score 197.8;  DB 9;  Length 163027;
 Best Local Similarity   89.8%;  Pred. No. 4e-28;
 Matches  214;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64

Db     108823 TAAAAATGGTTAAAATAGGCCGGGCACACTGGCTCACGCCTGTAATCCCAGCACTTCGGG 108882

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124

Db     108883 AGGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGT 108942

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184

Db     108943 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCGTGTGCCTGTAA 109002

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244

Db     109003 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGC 109062

Qy        245 A 245

Db     109063 A 109063


RESULT 12
AC096508/c
LOCUS       AC096508               163662 bp    DNA     linear   HTG 13-SEP-2002
DEFINITION  Homo sapiens chromosome X clone RP11-111F17, WORKING DRAFT
            SEQUENCE, 5 unordered pieces.
ACCESSION   AC096508
VERSION     AC096508.8  GI:22830224
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163662)
  AUTHORS   Muzny,D.M., Adams,C., Adic-Oducla,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,C., Foster,P., Frantz,F.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,C.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163662)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
```

```
JOURNAL     Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 163662)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 13, 2002 this sequence version replaced gi:22376857.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Drafting Center Code: BCM
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HDML
                Center clone name: RP11-111F17
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 161410 bases at least Q40
                Consensus quality: 162039 bases at least Q30
                Consensus quality: 162473 bases at least Q20
                Estimated insert size: 164107; sum-of-contigs estimation
                Quality coverage: 10x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     3227: contig of 3227 bp in length
            *     3228     3327: gap of unknown length
            *     3328    12653: contig of 9326 bp in length
            *    12654    12753: gap of unknown length
            *    12754    33426: contig of 20673 bp in length
            *    33427    33526: gap of unknown length
            *    33527    60991: contig of 27465 bp in length
            *    60992    61091: gap of unknown length
            *    61092   163662: contig of 102571 bp in length.
FEATURES             Location/Qualifiers
     source          1..163662
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-111F17"
BASE COUNT    43922 a  36511 c  38209 g  44615 t    405 others
ORIGIN

  Query Match             19.8%;  Score 197.8;  DB 2;  Length 163662;
  Best Local Similarity   88.8%;  Pred. No. 4e-28;
  Matches  214;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64
                |||||  ||  ||   ||||  ||||  |  ||||||| ||||||||||||||||  |
Db     155556 TAAAAATGGTTAAAATAGGCCGGGCACACTGGCTCACGCCTGTAATCCCAGCACTTCGGG 155497

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
              |||||||||||| |||||||||||||||| |||| ||||  ||| |||||||| |||  |
Db     155496 AGGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGT 155437

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184
              ||||||||||||||||||||||||||||||||||||||||  || ||| |||||||||||
Db     155436 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCGTGTGCCTGTAA 155377

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
              |||||||||||  ||||||||||  ||| |||||||||||||| ||||| | |||||||
Db     155376 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGC 155317

Qy        245 A 245
              |
Db     155316 A 155316


RESULT 13
AC068676/c
LOCUS       AC068676              191866 bp    DNA     linear   HTG 16-DEC-2000
DEFINITION  Homo sapiens clone RP11-768N8, WORKING DRAFT SEQUENCE, 14 unordered
            pieces.
ACCESSION   AC068676
VERSION     AC068676.3  GI:11875293
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 191866)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-768N8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 191866)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 16, 2000 this sequence version replaced gi:10440660.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10325
                Center clone name: 768_N_8
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 2% of reads
                Sequencing vector: Plasmid; n/a; 98% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 186020 bases at least Q40
                Consensus quality: 188219 bases at least Q30
                Consensus quality: 189180 bases at least Q20
                Insert size: 188000; agarose-fp
```

```
            Insert size: 190566; sum-of-contigs
            Quality coverage: 9.2 in Q20 bases; agarose-fp
            Quality coverage: 9.1 in Q20 ba.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 14 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *         1       527: contig of 527 bp in length
          *       528       627: gap of 100 bp
          *       628      1287: contig of 660 bp in length
          *      1288      1387: gap of 100 bp
          *      1388      1833: contig of 446 bp in length
          *      1834      1933: gap of 100 bp
          *      1934      2723: contig of 790 bp in length
          *      2724      2823: gap of 100 bp
          *      2824      4471: contig of 1648 bp in length
          *      4472      4571: gap of 100 bp
          *      4572      8184: contig of 3613 bp in length
          *      8185      8284: gap of 100 bp
          *      8285     18881: contig of 10597 bp in length
          *     18882     18981: gap of 100 bp
          *     18982     33829: contig of 14848 bp in length
          *     33830     33929: gap of 100 bp
          *     33930     51797: contig of 17868 bp in length
          *     51798     51897: gap of 100 bp
          *     51898     67219: contig of 15322 bp in length
          *     67220     67319: gap of 100 bp
          *     67320     88231: contig of 20912 bp in length
          *     88232     88331: gap of 100 bp
          *     88332    132689: contig of 44358 bp in length
          *    132690    132789: gap of 100 bp
          *    132790    170094: contig of 37305 bp in length
          *    170095    170194: gap of 100 bp
          *    170195    191866: contig of 21672 bp in length.
FEATURES             Location/Qualifiers
     source          1..191866
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-768N6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..527
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    628..1287
                     /note="assembly_fragment"
     misc_feature    1388..1833
                     /note="assembly_fragment"
     misc_feature    1934..2723
                     /note="assembly_fragment"
     misc_feature    2824..4471
                     /note="assembly_fragment"
     misc_feature    4572..8184
                     /note="assembly_fragment"
     misc_feature    8285..18881
                     /note="assembly_fragment"
     misc_feature    18982..33829
                     /note="assembly_fragment"
     misc_feature    33930..51797
                     /note="assembly_fragment"
     misc_feature    51898..67219
                     /note="assembly_fragment"
     misc_feature    67320..88231
                     /note="assembly_fragment"
     misc_feature    88332..132689
                     /note="assembly_fragment"
     misc_feature    132790..170094
                     /note="assembly_fragment"
     misc_feature    170195..191866
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    52298 a  42689 c  42735 g  52841 t   1303 others
ORIGIN

  Query Match             19.8%;  Score 197.8;  DB 2;  Length 191866;
  Best Local Similarity   85.8%;  Pred. No. 4e-28;
  Matches  214;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64
Db      33796 TAAAAATGGTTAAAATAGGCCGGGCACACTGGCTCACGCCTGTAATCCCAGCACTTCGGG 33737

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
Db      33736 AGGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGT 33677

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184
Db      33676 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCGTGTGCCTGTAA 33617

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
Db      33616 TCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGC 33557

Qy        245 A 245
              |
Db      33556 A 33556


RESULT 14
AC099804
LOCUS       AC099804               119922 bp    DNA     linear   PRI 28-AUG-2002
DEFINITION  Homo sapiens chromosome 17, clone CTD-2357A8, complete sequence.
ACCESSION   AC099804
VERSION     AC099804.3  GI:22532758
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 119922)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2357A8
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 119922)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
```

```
REFERENCE   3  (bases 1 to 119922)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,C.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 119922)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,C.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 28, 2002 this sequence version replaced gi:22981667.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10689
                Center clone name: 2357_A_8
            --------------.
FEATURES             Location/Qualifiers
     source          1..119922
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2357A8"
                     /clone_lib="CITD1 Human BAC"
     repeat_region   280..573
                     /rpt_family="AluSx"
     repeat_region   682..784
                     /rpt_family="L2"
     repeat_region   790..1531
                     /rpt_family="L1PA13"
     repeat_region   2246..2386
                     /rpt_family="MLT2B2"
     repeat_region   2752..2953
                     /rpt_family="AluY"
     repeat_region   3116..3248
                     /rpt_family="MIR3"
     repeat_region   complement(3357..3446)
                     /rpt_family="AluSp/q"
     repeat_region   complement(3447..3632)
                     /rpt_family="AluSx"
     repeat_region   complement(5336..5513)
                     /rpt_family="AluSg/x"
     repeat_region   complement(5738..6036)
                     /rpt_family="AluJb"
     repeat_region   6792..7153
                     /rpt_family="THE1B"
     repeat_region   7223..7425
                     /rpt_family="MSTA"
     repeat_region   7426..7721
                     /rpt_family="AluSq"
     repeat_region   7722..7905
                     /rpt_family="MSTA"
     repeat_region   complement(7906..8219)
                     /rpt_family="AluSq"
     repeat_region   8220..8289
                     /rpt_family="MSTA"
     repeat_region   8302..8483
                     /rpt_family="MER77"
     repeat_region   9281..9385
                     /rpt_family="HERV16"
     repeat_region   complement(9386..9520)
                     /rpt_family="FLAM_C"
     repeat_region   9521..9655
                     /rpt_family="HERV16"
     repeat_region   complement(10817..11516)
                     /rpt_family="MER50"
     repeat_region   12317..12337
                     /rpt_family="AT_rich"
     repeat_region   12488..13151
                     /rpt_family="L1MB6"
     repeat_region   13152..13456
                     /rpt_family="AluSx"
     repeat_region   13457..13674
                     /rpt_family="L1MB6"
     repeat_region   complement(13691..13832)
                     /rpt_family="AluJb"
     repeat_region   13875..14022
                     /rpt_family="L1MB5"
     repeat_region   16058..16097
                     /rpt_family="L1M4"
     repeat_region   16992..17013
                     /rpt_family="(A)n"
     repeat_region   17066..17462
                     /rpt_family="L1MB5"
     repeat_region   17464..17668
                     /rpt_family="MER2"
     repeat_region   17705..17806
                     /rpt_family="AluS"
     repeat_region   17819..17895
                     /rpt_family="MER2"
     repeat_region   complement(17916..18220)
                     /rpt_family="AluSx"
     repeat_region   18221..18367
                     /rpt_family="L1MB5"
     repeat_region   18811..18876
                     /rpt_family="L1MB3"
     repeat_region   18883..18953
                     /rpt_family="L1MB8"
     repeat_region   complement(18955..19262)
                     /rpt_family="AluY"
     unsure          18993..19040
                     /note="PCR product sequence only"
     repeat_region   19346..19752
                     /rpt_family="MLT1C"
```

```
    unsure          19361
                    /note="Probably G; possibly T"
    unsure          19412..19417
                    /note="<30 qual SNGL region"
    unsure          19459..19530
                    /note="single clone coverage"
    repeat_region   20040..20337
                    /rpt_family="AluSp"
    unsure          20341..20345
                    /note="<30 qual SNGL region"
    unsure          20425..20443
                    /note="<30 qual SNGL region"
    repeat_region   20450..20745
                    /rpt_family="AluJo"
    unsure          20495..20503

  Query Match             19.7%;  Score 196.8;  DB 9;  Length 119922;
  Best Local Similarity   88.8%;  Pred. No. 6.4e-28;
  Matches  213;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db       7410 AAAAGTGTGTAGCACTGGCTGGGCACAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 7469

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db       7470 GGCCGAGGTGGGCAGATCACTCGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCG 7529

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db       7530 AAACCCTGTCTCCACTAAAAATACAAAAATTAGCTGGGCATGGTGGTGGGTGCCTGTAAT 7589

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db       7590 CCCAGCTACTCAGGAGGCTGAGGCAGAAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCA 7649


RESULT 15
HSJ738A13/c
LOCUS       HSJ738A13               46775 bp    DNA     linear   PRI 01-FEB-2003
DEFINITION  Human DNA sequence from clone RP4-738A13 on chromosome
            Xq21.31-22.1, complete sequence.
ACCESSION   AL109801
VERSION     AL109801.14  GI:28193390
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 46775)
  AUTHORS   Whitehead,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2003 this sequence version replaced gi:6624938.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            RP4-738A13 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..46775
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q21.31-22.1"
                     /clone="RP4-738A13"
                     /clone_lib="RPCI-4"
BASE COUNT    13491 a   9466 c   9528 g  14290 t
ORIGIN

  Query Match             19.6%;  Score 195.8;  DB 9;  Length 46775;
  Best Local Similarity   88.7%;  Pred. No. 1.1e-27;
  Matches  212;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          7 AAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAG 66
Db      39786 AAAAGTATAAATTTAGTCCGGGCACGGTGGCTCACTCCTGTAATCCCAGCACTTTGGGAG 39727

Qy         67 GCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGA 126
Db      39726 GCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGATGA 39667

Qy        127 AACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATC 186
Db      39666 AACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCATGTGCCTGTAATC 39607

Qy        187 CCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db      39606 CCAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCCAGGAGGTGGAGGTTGCA 39548


Search completed: October 24, 2003, 21:06:17
Job time : 2773.61 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 252.249 Seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....aactgattctggtctttattt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1001	100.0	110000	22	AAF84800	Nucleotide sequenc
2	403	40.3	3263	22	AAF84801	Nucleotide sequenc
3	403	40.3	4152	24	ABQ54772	Human ovarian anti
4	191.4	19.1	3143	24	AAS17746	Human genomic clon
5	191.4	19.1	30573	22	AAL04181	Human reproductive
6	191.4	19.1	32195	22	AAL04180	Human reproductive
7	190.8	19.1	7677	23	ABV23205	Human prostate exp
8	190.8	19.1	7677	23	ABV29042	Human prostate exp
c 9	190.4	19.0	3074	21	AAC76994	Human CRFX ORF2549
10	189.8	19.0	10380	18	AAT67164	Human alpha-N-acet
11	188.8	18.9	17792	22	AAS32727	Human genomic DNA
12	188.8	18.9	17792	22	AAS36099	Human cardiovascul
c 13	188.6	18.8	19616	22	ABA16094	Human nervous syst
c 14	188.6	18.8	19616	22	ABA16856	Human nervous syst
c 15	188.6	18.8	19616	22	AAL36855	Human musculoskele
c 16	188.6	18.8	19616	25	ABX59843	cDNA encoding nove
17	187.8	18.8	50000	21	AAA96366	Polymorphic repeat
c 18	187.6	18.7	4768	21	AAC69134	Human ABC1 gene ex
c 19	187.6	18.7	11754	22	AAS04033	Human ABC1 gene, p
c 20	187.6	18.7	11754	24	ABL58149	Human ABCA1 transp
c 21	187.6	18.7	52616	22	AAK70459	Human immune/haema
c 22	187.6	18.7	52616	22	AAK78930	Human immune/haema
c 23	187.6	18.7	97835	24	ABK84796	Human cDNA differe
c 24	187.6	18.7	183999	22	AAF92831	Human ABC1 genomic
25	187	18.7	3417	22	AAH18467	Human cDNA sequenc
26	186.6	18.6	812	21	AAC74416	Human secreted pro
c 27	186.6	18.6	6616	22	AAK81014	Human immune/haema
28	186.6	18.6	9469	22	AAK79514	Human immune/haema
29	186.6	18.6	17335	23	ABK42393	Genomic sequence #
30	186.6	18.6	19882	23	ABK42394	Genomic sequence #
31	186.6	18.6	79528	24	AAL50614	Human cancer statu
c 32	186.6	18.6	116840	25	ABQ77401	Human VWF DNA. Ho
33	186.6	18.6	1691080	24	ABX08336	Human phosphodiest
34	186.4	18.6	2233	22	AAK77256	Human immune/haema
c 35	186.4	18.6	13467	22	AAL05944	Human reproductive
c 36	186.4	18.6	13467	22	AAS27670	DNA encoding novel
c 37	186.4	18.6	13467	22	AAS27838	DNA encoding novel
c 38	186.4	18.6	13467	22	AAK85539	Human immune/haema
c 39	186.4	18.6	13467	23	ABL98508	Human testicular a
c 40	186.2	18.6	14491	25	ABZ09914	Human 5' and/or re
c 41	185.8	18.6	32190	22	AAS36709	Human cardiovascul
c 42	185.6	18.5	13058	22	AAS29194	Genomic sequence #
43	185.6	18.5	13058	22	AAK79325	Human immune/haema
c 44	185.6	18.5	13058	24	ABS68334	Human DNA-binding
c 45	185.6	18.5	62804	24	AAD39317	Human calcium/calm

ALIGNMENTS

```
RESULT 1
AAF84800
ID   AAF84800 standard; DNA; 110000 BP.
XX
AC   AAF84800;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the human SPG4 gene.
XX
KW   Human; SPG4 gene; spastin, PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9932..102009
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            9932..10471
FT                   /*tag= b
FT                   /number= "1"
FT   intron          10472..33718
FT                   /*tag= c
FT                   /number= "1"
FT   exon            33719..33805
FT                   /*tag= d
FT                   /number= "2"
FT   intron          33806..35748
FT                   /*tag= e
FT                   /number= "2"
```

```
FT   exon            35749..35832
FT                   /*tag= f
FT                   /number= "3"
FT   intron          35833..45022
FT                   /*tag= g
FT                   /number= "3"
FT   exon            45023..45118
FT                   /*tag= h
FT                   /number= "4"
FT   intron          45119..60863
FT                   /*tag= i
FT                   /number= "4"
FT   exon            60864..61051
FT                   /*tag= j
FT                   /number= "5"
FT   intron          61052..61927
FT                   /*tag= k
FT                   /number= "5"
FT   exon            61928..62061
FT                   /*tag= l
FT                   /number= "6"
FT   intron          62062..62344
FT                   /*tag= m
FT                   /number= "6"
FT   exon            62345..62438
FT                   /*tag= n
FT                   /number= "7"
FT   intron          62439..73173
FT                   /*tag= o
FT                   /number= "7"
FT   exon            73174..73248
FT                   /*tag= p
FT                   /number= "8"
FT   intron          73249..74633
FT                   /*tag= q
FT                   /number= "8"
FT   exon            74634..74705
FT                   /*tag= r
FT                   /number= "9"
FT   intron          74706..82788
FT                   /*tag= s
FT                   /number= "9"
FT   exon            82789..82864
FT                   /*tag= t
FT                   /number= "10"
FT   intron          82865..83102
FT                   /*tag= u
FT                   /number= "10"
FT   exon            83103..83194
FT                   /*tag= v
FT                   /number= "11"
FT   intron          83195..83334
FT                   /*tag= w
FT                   /number= "11"
FT   exon            83335..83414
FT                   /*tag= x
FT                   /number= "12"
FT   intron          83415..88129
FT                   /*tag= y
FT                   /number= "12"
FT   exon            88130..88172
FT                   /*tag= z
FT                   /number= "13"
FT   intron          88173..89561
FT                   /*tag= aa
FT                   /number= "13"
FT   exon            89562..89641
FT                   /*tag= ab
FT                   /number= "14"
FT   intron          89642..91162
FT                   /*tag= ac
FT                   /number= "14"
FT   exon            91163..91233
FT                   /*tag= ad
FT                   /number= "15"
FT   intron          91234..93443
FT                   /*tag= ae
FT                   /number= "15"
FT   exon            93444..93484
FT                   /*tag= af
FT                   /number= "16"
FT   intron          93485..100599
FT                   /*tag= ag
FT                   /number= "16"
FT   exon            100600..102009
FT                   /*tag= ah
FT                   /number= "17"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP 1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J, Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening -
XX
PS   Claim 2; Page 45-106; 146pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 other;

  Query Match             100.0%;  Score 1001;  DB 22;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2e-171;
  Matches 1001;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100000 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 100059

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100060 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 100119

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100120 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 100179

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100180 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 100239

Qy        241 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     100240 TTGCATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTGAACT 100299

Qy        301 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   100300 TTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTA 100359

Qy      361 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 420
Db   100360 ATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGTTTCT 100419

Qy      421 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 480
Db   100420 TGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAATAC 100479

Qy      481 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 540
Db   100480 ATACACGTATATTTTTTATAACATTAAGAAACAGCAGCATCATTACTTTAATCCATCATT 100539

Qy      541 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 600
Db   100540 TCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAAAAATCTAG 100599

Qy      601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 660
Db   100600 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTC 100659

Qy      661 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 720
Db   100660 AGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTT 100719

Qy      721 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 780
Db   100720 TAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCT 100779

Qy      781 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 840
Db   100780 TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTAC 100839

Qy      841 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 900
Db   100840 ATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATG 100899

Qy      901 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 960
Db   100900 CAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGAC 100959

Qy      961 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db   100960 ACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 101000
```

```
RESULT 2
AAF84801
ID   AAF84801 standard; cDNA; 3263 BP.
XX
AC   AAF84801;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the human SPG4 polypeptide.
XX
KW   Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             126..1976
FT                   /*tag=  a
FT                   /product= "spastin"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
DR   P-PSDB; AAB68137.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening  -
XX
PS   Claim 2; Page 106-108; 145pp; French.
XX
CC   The present sequence encodes a human spastin polypeptide. Spastin is
CC   encoded by the SPG4 gene. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 3263 BP; 956 A; 664 C; 703 G; 940 T; 0 other;

  Query Match             40.3%;  Score 403;  DB 22;  Length 3263;
  Best Local Similarity   100.0%;  Pred. No. 8e-64;
  Matches  403;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
Db       1852 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 1911

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
Db       1912 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 1971

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
Db       1972 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 2031

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
Db       2032 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 2091

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
Db       2092 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 2151

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
Db       2152 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 2211

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db       2212 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 2254
```

```
RESULT 3
ABQ54772
ID   ABQ54772 standard; cDNA; 4152 BP.
XX
AC   ABQ54772;
XX
DT   22-AUG-2002  (first entry)
XX
DE   Human ovarian antigen HNTTB23 cDNA, SEQ ID NO:652.
XX
KW   Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW   ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW   infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
```

KW PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW antiinflammatory; gynaecological; reproductive; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WO-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-147878/19.
DR P-PSDB; ABP41695.
XX
PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; SEQ ID No 652; 2922pp; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4152 BP; 1331 A; 663 C; 777 G; 1376 T; 5 other;

Query Match 40.3%; Score 403; DB 24; Length 4152;
Best Local Similarity 100.0%; Pred. No. 8.1e-64;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
Db 853 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 912

Qy 659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
Db 913 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 972

Qy 719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
Db 973 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 1032

Qy 779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
Db 1033 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 1092

Qy 839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
Db 1093 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 1152

Qy 899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
Db 1153 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 1212

Qy 959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db 1213 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1255

RESULT 4
AAS17746
ID AAS17746 standard; DNA; 3143 BP.
XX
AC AAS17746;
XX
DT 26-FEB-2002 (first entry)
XX
DE Human genomic clone for P2Y-like G protein-coupled receptor.
XX
KW Human; ds; P2Y-like G protein-coupled receptor; GPCR; COPD;
KW chronic obstructive pulmonary disease; nervous system disease;
KW Parkinson's disease; multiple sclerosis; dementia; stroke;
KW Alzheimer's disease; benign prostatic hyperplasia; urinary incontinence;
KW bacterial infection; fungal infection; protozoan infection;
KW viral infection; pain; cancer; anorexia; bulimia; asthma;
KW acute heart failure; hypotension; hypertension; osteoporosis; diabetes;
KW angina pectoris; myocardial infarction; ulcer; inflammation; allergy;
KW psychotic disorder; neurological disorder; anxiety; schizophrenia;
KW manic depression; delirium; severe mental retardation; dyskinesia.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 520..2498
FT /*tag= a
FT /product= "P2Y-like GPCR"
XX
PN WO200185764-A2.
XX
PD 15-NOV-2001.
XX
PF 09-MAY-2001; 2001WO-EP05244.
XX
PR 11-MAY-2000; 2000US-203582P.
PR 21-FEB-2001; 2001US-269857P.
XX
PA (FARB) BAYER AG.
XX
PI Ramakrishnan S;
XX
DR WPI; 2002-075242/10.
DR P-PSDB; AAU11251.
XX
PT New polynucleotides for producing P2Y-like G protein-coupled receptors
PT (GPCR) that are used for screening inhibitors or regulators of human
PT P2Y-like GPCR, especially useful for treating pain, cancer or
PT neurological disorders -

```
XX
PS   Disclosure; Fig 1; 114pp; English.
XX
CC   The invention relates to an isolated polynucleotide encoding a P2Y-like
CC   G protein-coupled receptor (GPCR) polypeptide, its fragment,
CC   derivative or allele, a host cell containing an expression vector
CC   comprising the polynucleotide and screening for agents that regulate the
CC   GPCR activity. The polynucleotide is useful for producing P2Y-like GPCR
CC   polypeptide, which may be employed for screening agents that inhibit or
CC   regulate human P2Y-like GPCR. The reagent or inhibitor of the human
CC   P2Y-like GPCR is useful for treating or ameliorating P2Y-like GPCR
CC   disorders, particularly COPD (chronic obstructive pulmonary disease),
CC   peripheral or central nervous system disease (e.g. Parkinson's
CC   disease, multiple sclerosis, dementia, stroke, Alzheimer's disease and
CC   many other diseases and disorders listed in the specification), benign
CC   prostatic hyperplasia or urinary incontinence. A pharmaceutical
CC   composition containing the modulators and/or regulators of P2Y-like
CC   GPCR is useful for modulating the activity of a P2Y-like GPCR.
CC   In particular, these are useful for treating, preventing or ameliorating
CC   infections (e.g. bacterial, fungal, protozoan or viral infections), pain,
CC   cancer, anorexia, bulimia, asthma, acute heart failure, hypotension,
CC   hypertension, osteoporosis, diabetes, angina pectoris, myocardial
CC   infarction, ulcers, inflammation, allergies, psychotic or neurological
CC   disorders (e.g. anxiety, schizophrenia, manic depression, delirium,
CC   severe mental retardation or dyskinesias). The present sequence is
CC   a genomic clone encoding the P2Y-like GPCR of the invention.
XX
SQ   Sequence 3143 BP; 818 A; 771 C; 726 G; 828 T; 0 other;

  Query Match             19.1%;  Score 191.4;  DB 24;  Length 3143;
  Best Local Similarity   88.8%;  Pred. No. 9.9e-26;
  Matches  207;  Conservative    0;  Mismatches   26;  Indels    3;  Gaps    0;

Qy         13 CATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAG 72
Db       2723 CATTATCTAGGCTGGGCACAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 2782

Qy         73 GTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCT 132
Db       2783 GCGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAATCTGGCCAACATGGTGAAACCCC 2842

Qy        133 GTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCT 192
Db       2843 ATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCGGGTGCCTATAATCCCAGCT 2902

Qy        193 ACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db       2903 ACTCGGGAGGCTGAGAAAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCA 2955


RESULT 5
AAL04181
ID   AAL04181 standard; DNA; 30573 BP.
XX
AC   AAL04181;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 6969.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01339.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   26-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
```

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6869; 1297pp + Sequence Listing; English.
XX

CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 30573 BP; 7764 A; 7219 C; 7253 G; 8337 T; 0 other;

Query Match 19.1%; Score 191.4; DB 22; Length 30573;
Best Local Similarity 90.7%; Pred. No. 1.1e-25;
Matches 204; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 21 AGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGG 80
Db 6984 AGGCTGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGTCGAGGTAGGTGG 7043

Qy 81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTAC 140
Db 7044 ATCATCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 7103

Qy 141 TAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGA 200
Db 7104 TAAAAATACAACAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTTGGGA 7163

Qy 201 GGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db 7164 GGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCA 7208

RESULT 6
AAL04180
ID AAL04180 standard; DNA; 32195 BP.
XX
AC AAL04180;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6868.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 28-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-465570/50.
XX
PT Isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition -
XX
PS Disclosure; SEQ ID NO 6868; 1297pp - Sequence Listing; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 32195 BP; 8257 A; 7420 C; 7374 G; 9144 T; 0 other;

Query Match 19.1%; Score 191.4; DB 22; Length 32195;
Best Local Similarity 90.7%; Pred. No. 1.1e-25;
Matches 204; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

Qy 21 AGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGG 80
Db 12766 AGGCTGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGTCGAGGTAGGTGG 12825

Qy 81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTAC 140
Db 12826 ATCATCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 12885

Qy 141 TAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGA 200

```
Db      12886 TAAAAATACAACAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTTGGGA 12945

Qy        201 GGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      12946 GGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCA 12990
```

RESULT 7
ABV23205
ID ABV23205 standard; cDNA; 7677 BP.
XX
AC ABV23205;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 23196.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 4173-4175; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ Sequence 7677 BP; 2389 A; 1642 C; 1794 G; 1846 T; 6 other;

Query Match 19.1%; Score 190.8; DB 23; Length 7677;
Best Local Similarity 60.7%; Pred. No. 1.3e-25;
Matches 368; Conservative 0; Mismatches 227; Indels 11; Gaps 3;

```
Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60

Db       6836 AATTTATTAAGAATTAATCTCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTT 6895

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120

Db       6896 TGGGAAGCCAAGGTGGGCGGATCACTTCAGGTTAGGAGTTGGAGACCAGCCTGGCCAACA 6955

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180

Db       6956 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCACTTGCCT 7015

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240

Db       7016 GTAATCCCAGCTACTCGGGAGGCTGAGGCACGAGAATCGATTAAACCTGGGAGGCGGAGA 7075

Qy        241 TTGC----ATATCTGAGTGGTGAAATTGTGATTCTTTTTTCTCTTTGTCTGTATTTTG 296

Db       7076 TTGCAATGAGATCGAACCACTGCACTCCAGCCTGGGTGACAGAGAGAGACACTGCCTTGG 7135

Qy        297 AACTTTTCTATAAATGATTGTGTTTTGTTTTATATTGGAAAAATATTATGCTTTCAAAT 356

Db       7136 AAAAAAAAAGAATCTCACTCACTATCTAGAGAGGATTGTCAGAATATTCACGATTCAGGT 7195

Qy        357 GTTAATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGT 416

Db       7196 CTTGAAACTTTGATTATGCAAAAGAAGGTATATAATAAATATTTCATTATGATTCAGTTT 7255

Qy        417 TTCTTGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGA 476

Db       7256 TTAAGGCTTTGCAGCT---TCTATAAGTGTTCTCAGATGCCACTAGATAATTTTAAAAGC 7312

Qy        477 ATACATACACGTATATTTTTATAACATTA----AGAAACAGCAGCATCATTACTTTAAT 532

Db       7313 ATCATATTAGAAATACTTTAAGAAGACTTATATAAGAAATAGAAGATTGTTGAATTTTAC 7372

Qy        533 CCATCATTTCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAA 592

Db       7373 AGAGGATTTGGTTCATTAAGACCCAGATTCTGTAAGTTTTCATTCTGAAATTCTAGTTAA 7432

Qy        593 AAATCT 598

Db       7433 ACATAT 7438
```

RESULT 8
ABV29042
ID ABV29042 standard; cDNA; 7677 BP.
XX
AC ABV29042;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 29033.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US05171.
XX
PR 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;

```
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS   Claim 1; Page 6152-6154; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 7677 BP; 2389 A; 1642 C; 1794 G; 1846 T; 6 other;

  Query Match             19.1%;  Score 190.8;  DB 23;  Length 7677;
  Best Local Similarity   60.7%;  Pred. No. 1.3e-25;
  Matches  368;  Conservative    0;  Mismatches  227;  Indels   11;  Gaps    3;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60
Db       6836 AATTTATTAAGAATTAATCTCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTT 6895

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
Db       6896 TGGGAAGCCAAGGTGGGCGGATCACTTCAGGTTAGGAGTTGGAGACCAGCCTGGCCAACA 6955

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180
Db       6956 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCACTTGCCT 7015

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240
Db       7016 GTAATCCCAGCTACTCGGGAGGCTGAGGCACGAGAATCGATTAAACCTGGGAGGCGGAGA 7075

Qy        241 TTGC----ATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTG 296
Db       7076 TTGCAATGAGATCGAACCACTGCACTCCAGCCTGGGTGACAGAGAGAGACACTGCCTTGG 7135

Qy        297 AACTTTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAAT 356
Db       7136 AAAAAAAAAGAATCTCACTCACTATCTAGAGAGGATTGTCAGAATATTCACGATTCAGGT 7195

Qy        357 GTTAATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGT 416
Db       7196 CTTGAAACTTTGATTATGCAAAAGAAGGTATATAATAAATATTTCATTATGATTCAGTTT 7255

Qy        417 TTCTTGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGA 476
Db       7256 TTAAGGCTTTGCAGCT---TCTATAAGTGTTCTCAGATGCCACTAGATAATTTTAAAAGC 7312

Qy        477 ATACATACACGTATATTTTTTATAACATTA----AGAAACAGCAGCATCATTACTTTAAT 532
Db       7313 ATCATATTAGAAATACTTTAAGAAGACTTATATAAGAAATAGAAGATTGTTGAATTTTAC 7372

Qy        533 CCATCATTTCGTTAACCACCATATACCTGTTGATCATTTGTATTGTCATGTGCTTTTTAA 592
Db       7373 AGAGGATTTGGTTCATTAAGACCCAGATTCTGTAAGTTTTCATTCTGAAATTCTAGTTAA 7432

Qy        593 AAATCT 598
Db       7433 ACATAT 7438


RESULT 9
AAC76994/c
ID   AAC76994 standard; cDNA; 3074 BP.
XX
AC   AAC76994;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human ORFX ORF2549 polynucleotide sequence SEQ ID NO:5097.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB42785.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 4285-4287; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
```

```
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 3074 BP; 654 A; 851 C; 847 G; 720 T; 2 other;

  Query Match             19.0%;  Score 190.4;  DB 21;  Length 3074;
  Best Local Similarity   87.1%;  Pred. No. 1.5e-25;
  Matches  209;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db       1593 AAAAGTGCCTTTTAGAGGCTGGGCATGGTGGCTCACGCTTGTAATCCCAGCACTTTGGGA 1534

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db       1533 GGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTG 1474

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db       1473 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTAGCAGGTGCCTGTAAT 1414

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db       1413 CCCAGCTACTGGGGAAGCTGAGGCAGGAGAATTGCTTCAATCTGGAAGGCAGAGGTTGCA 1354


RESULT 10
AAT67164
ID   AAT67164 standard; cDNA; 10380 BP.
XX
AC   AAT67164;
XX
DT   20-AUG-1997  (first entry)
XX
DE   Human alpha-N-acetylglucosaminidase gene.
XX
KW   Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW   gene therapy; enzyme replacement therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..989
FT                   /*tag= a
FT   exon            990..1372
FT                   /*tag= b
FT                   /number= 1
FT   intron          1373..2114
FT                   /*tag= c
FT   exon            2115..2262
FT                   /*tag= d
FT                   /number= 2
FT   intron          2263..3055
FT                   /*tag= e
FT   intron          3056..3202
FT                   /*tag= f
FT                   /number= 3
FT   intron          3203..3386
FT                   /*tag= g
FT   exon            3387..3472
FT                   /*tag= h
FT                   /number= 4
FT   intron          3473..5666
FT                   /*tag= i
FT   exon            5667..5923
FT                   /*tag= j
FT                   /number= 5
FT   intron          5924..7744
FT                   /*tag= k
FT   exon            7745..8955
FT                   /*tag= l
FT                   /number= 6
FT   3'UTR           8966..10380
FT                   /*tag= m
XX
PN   WO9719177-A1.
XX
PD   29-MAY-1997.
XX
PF   22-NOV-1996;   96WO-AU00747.
XX
PR   23-NOV-1995;   95AU-0006749.
XX
PA   (WOME-) WOMEN'S & CHILDREN'S HOSPITAL.
XX
PI   Anson DS,  Blanch L,  Hopwood JJ,  Scott H,  Weber B;
XX
DR   WPI; 1997-298114/27.
DR   P-PSDB; AAW18017.
XX
PT   Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -
PT   used for the diagnosis and treatment of mucopolysaccharidosis type
PT   IIIB; also used in gene therapy
XX
PS   Claim 8; Page 54-61; 79pp; English.
XX
CC   A genomic DNA molecule (AAT67164) includes 6 exons that code for
CC   human alpha-N-acetylglucosaminidase (AAW18017), an enzyme that can
CC   hydrolyse the terminal alpha-N-acetylglucosamine residues at the
CC   non-reducing terminus of fragments of heparan sulphate and heparin.
CC   It was isolated by hybridisation of a human chromosome 17 library.
CC   A cDNA clone (AAT67163) coding for the enzyme has also been isolated.
CC   The isolated gene or cDNA, and primers/probes based on them or
CC   their complementary strands, can be used to investigate, diagnose
CC   and treat alpha-N-acetylglucosaminidase deficiency, for example in
CC   patients suffering from mucopolysaccharidosis type IIIB.
CC   Administration is by oral, i.v., i.p., enzyme replacement therapy,
CC   gene therapy or other routes.
XX
SQ   Sequence 10380 BP; 2210 A; 2953 C; 2851 G; 2366 T; 0 other;

  Query Match             19.0%;  Score 189.8;  DB 18;  Length 10380;
  Best Local Similarity   86.7%;  Pred. No. 2e-25;
  Matches  209;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64
Db       3540 TAAAAACTTAAGCTCTGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 3599

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
Db       3600 AGGCCGAGTTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGT 3659

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184
Db       3660 GAAACCCTGTCTCTACTAAAAATATAAAAATTAGCTGGGCATGGTGGCACGCGCCTGTAA 3719

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
Db       3720 TCCCATCTACTTAGGAGGCTGAGACAGGAGAATTGCTTAAACCTGGGAGGCAGACGTTGC 3779

Qy        245 A 245

Db       3780 A 3780


RESULT 11
AAS32727
ID   AAS32727 standard; DNA; 17792 BP.
XX
AC   AAS32727;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human genomic DNA for novel endocrine antigen, SEQ ID No 681.
XX
```

KW Human; endocrine antigen; ds; cytostatic; antiinfertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric; gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome.
XX
OS Homo sapiens.
XX
PN WO200155319-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01335.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0238935.
PR 13-OCT-2000; 2000US-0238937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.

```
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-457726/49.
XX
PT   Isolated polypeptide for treating, preventing and prognosing disorders
PT   related to the endocrine system including endocrine disorders,
PT   reproductive disorders, and gastrointestinal disorders and also for
PT   testing and detection e.g. diagnosis -
XX
PS   Disclosure; SEQ ID No 681; 558pp; English.
XX
CC   The invention relates to cDNAs encoding novel human endocrine
CC   antigens or a fragment having biological activity, a domain, an epitope,
CC   full length protein, variant, allelic variant or a species homologue of
CC   the cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition when administered
CC   (e.g. by gene therapy or antisense-therapy). Identifying mutations in
CC   the genes coding for the antigens is useful for diagnosing a pathological
CC   condition or a susceptibility to a pathological condition. The DNAs,
CC   antigens and antibodies raised against the antigens useful for treating,
CC   preventing and/ or prognosing disorders related to the endocrine system
CC   or hormone imbalance or reproductive disorders, cancers of endocrine
CC   tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal
CC   glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC   hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC   of diseases and disorders are given in the specification. The present
CC   sequence is genomic DNA fragment form a gene encoding an endocrine
CC   antigen of the invention.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 17792 BP; 4372 A; 4626 C; 4413 G; 4381 T; 0 other;

  Query Match             18.9%;  Score 186.8;  DB 22;  Length 17792;
  Best Local Similarity   86.7%;  Pred. No. 3.1e-25;
  Matches  208;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
                |||||  ||  |||||||||||  |||||||||  |||||||||||||||||   |||
Db      13401 AAAAAATCACTCTTTCGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 13460

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
              |||||||||||||||||||||||||| ||| ||||| |     |||  |||  || |||  |
Db      13461 GGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 13520

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
              ||||||  |||||||||||||||||||||||||||| |||   |||   ||      |||||
Db      13521 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCACACACCTGTAAT 13580

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
              |||||||||| |||||||||||  ||  | |||  |||||||  |||||  |  |||||||
Db      13581 CCCAGCTACTTGGGAGGCTGAGGCAAGAGAATAGCTTGAACCTGGGAGATGGAGGTTGCA 13640


RESULT 12
AAS36099
ID   AAS36099 standard; DNA; 17792 BP.
XX
AC   AAS36099;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Human cardiovascular system antigen genomic DNA SEQ ID No 1599.
XX
KW   Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW   chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; tissue regeneration;
KW   anti-infertility.
XX
OS   Homo sapiens.
XX
PN   WO200155321-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01340.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
```

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-451930/48.
XX
PT New cardiovascular system related polynucleotides and polypeptides,
PT useful for diagnosing, treating and/or preventing disorders of the
PT cardiovascular system -
XX
PS Claim 1; SEQ ID No 1599; 674pp; English.
XX
CC Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
CC the cardiovascular system antigen polypeptides of the invention.
CC Cardiovascular system antigens and their associated polynucleotides are
CC useful in the diagnosis, treatment and prevention of various types of
CC disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
CC chickens or sheep. A pathological condition can be determined by
CC detecting the presence or absence of a mutation in a cardiovascular
CC system antigen polynucleotide. The treatable disorders include autoimmune
CC diseases such as rheumatoid arthritis, hyperproliferative disorders such
CC as neoplasms of the breast or liver, cardiovascular disorders such as
CC cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
CC nervous system disorders such as Alzheimer's disease, infections caused
CC by bacteria, viruses and fungi, ocular disorders such as corneal
CC infection, endocrine disorders such as premature labour and infertility,
CC gastrointestinal disorders such as Crohn's disease, renal disorders such
CC as glomerulonephritis and respiratory disorders such as asthma and
CC pleurisy. The polypeptides can also be used to aid wound healing, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, to regenerate tissues and in chemotaxis.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

Query Match 18.9%; Score 188.8; DB 22; Length 19792;
Best Local Similarity 86.7%; Pred. No. 3.1e-25;
Matches 208; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db 13401 AAAAAATCACTCTTTCGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 13460

Qy 66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db 13461 GGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 13520

Qy 126 AAACCCTGTCTCTACTAAAAATACAAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db 13521 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCACACACCTGTAAT 13580

Qy 186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db 13581 CCCAGCTACTTGGGAGGCTGAGGCAAGAGAATAGCTTGAACCTGGGAGATGGAGGTTGCA 13640

RESULT 13
ABA16094/c
ID ABA16094 standard; DNA; 19616 BP.
XX

AC ABA16094;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8425.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180626.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216980.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241805.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.

```
PR   01-DEC-2000; 2000US-0251160.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-541565/60.
XX
PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating nervous system
PT   cancers and metastases -
XX
PS   Disclosure; SEQ ID NO 8425; 1701pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 19616 BP; 4181 A; 5695 C; 5539 G; 4201 T; 0 other;

  Query Match             18.8%;  Score 188.6;  DB 22;  Length 19616;
  Best Local Similarity   86.0%;  Pred. No. 3.4e-25;
  Matches  209;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62

Db      17028 TGTGTAAAATGCTAAATGGGGCCAAGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTG 16969

Qy         63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122

Db      16968 GGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATG 16909

Qy        123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182

Db      16908 ACGAAACCCTGTCTCTACTAAAAATACAAAATTTAGCTGGGCGTGGTGGCACATGCCTGT 16849

Qy        183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242

Db      16848 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAAGCGGAGGTT 16789

Qy        243 GCA 245

Db      16788 GCA 16786


RESULT 14
ABA16856/c
ID   ABA16856 standard; DNA; 19616 BP.
XX
AC   ABA16856;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 9187.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   23-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
```

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 9187; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 19616 BP; 4181 A; 5695 C; 5539 G; 4201 T; 0 other;

Query Match 18.8%; Score 188.6; DB 22; Length 19616;
Best Local Similarity 86.0%; Pred. No. 3.4e-25;
Matches 209; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62
Db 17028 TGTGTAAAATGCTAAATGGGGCCAAGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTG 16969

Qy 63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122
Db 16968 GGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATG 16909

Qy 123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182
Db 16908 ACGAAACCCTGTCTCTACTAAAAATACAAAATTTAGCTGGGCGTGGTGGCACATGCCTGT 16849

Qy 183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242
Db 16848 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAAGCGGAGGTT 16789

Qy 243 GCA 245
Db 16788 GCA 16786

```
RESULT 15
AAL36855/c
ID   AAL36855 standard; DNA; 19616 BP.
XX
AC   AAL36855;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 3220.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226869.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   28-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
```

```
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451937/48.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the musculoskeletal system including
PT   musculoskeletal cancers and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Example 2; SEQ ID NO 3220; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 19616 BP; 4181 A; 5695 C; 5539 G; 4201 T; 0 other;

  Query Match             18.8%;  Score 188.6;  DB 22;  Length 19616;
  Best Local Similarity   86.0%;  Pred. No. 3.4e-25;
  Matches  209;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62

Db      17028 TGTGTAAAATGCTAAATGGGGCCAAGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTG 16969

Qy         63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122

Db      16968 GGAGGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATG 16909

Qy        123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182

Db      16908 ACGAAACCCTGTCTCTACTAAAAATACAAAATTTAGCTGGGCGTGGTGGCACATGCCTGT 16849

Qy        183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242

Db      16848 AATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGAAGCGGAGGTT 16789

Qy        243 GCA 245

Db      16788 GCA 16786


Search completed: October 24, 2003, 18:47:25
Job time : 255.249 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 13:34:43 ; Search time 62.6458 Seconds
(without alignments)
7052.739 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....acctgattctggctttctttt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220681566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		189.8	19.0	10380	3	US-09-077-354B-3	Sequence 3, Appl
2		188.8	18.9	98844	4	US-09-791-211-10	Sequence 10, Appl
3	c	185.6	18.5	62804	4	US-09-800-960-3	Sequence 3, Appli
4		184	18.4	283	4	US-08-579-445-26	Sequence 26, Appl
5		183.6	18.3	282	1	US-08-133-629-8	Sequence 8, Appli
6		183.6	18.3	1154	4	US-09-539-333D-37	Sequence 37, Appl
7		183.6	18.3	1301	4	US-09-539-333D-36	Sequence 36, Appl
8		183.6	18.3	1386	4	US-09-539-333D-40	Sequence 40, Appl
9		183.6	18.3	319608	4	US-09-679-409-1	Sequence 1, Appli
10		183.2	18.3	319608	4	US-09-539-333D-1	Sequence 1, Appli
11	c	182.4	18.2	81001	4	US-09-750-580-1	Sequence 1, Appli
12		181.8	18.2	23187	4	US-09-499-522-1	Sequence 1, Appli
13	c	180.6	18.0	21784	4	US-09-820-002-3	Sequence 3, Appli
14		180.2	18.0	1762	4	US-09-443-184-35	Sequence 35, Appl
15	c	180	18.0	55298	4	US-09-491-356C-1	Sequence 1, Appli
16	c	179.8	18.0	50000	4	US-09-146-053-3	Sequence 3, Appli
17	c	179.4	17.9	3035	1	US-08-726-725-2	Sequence 2, Appli
18		179	17.9	11811	3	US-09-078-294-7	Sequence 7, Appli
19		179	17.9	29629	4	US-09-729-995-3	Sequence 3, Appli
20		179	17.9	62804	4	US-09-800-960-3	Sequence 3, Appli
21	c	178.6	17.8	1019	4	US-09-177-650-128	Sequence 128, App
22	c	178.6	17.8	12597	4	US-09-705-289-12	Sequence 12, Appl
23		177.6	17.7	45546	4	US-09-146-053-6	Sequence 6, Appli
24		177.4	17.7	2000	4	US-09-705-267A-19	Sequence 19, Appl
25		177.4	17.7	18000	4	US-09-657-346A-17	Sequence 17, Appl
26	c	177.2	17.7	92139	4	US-09-918-686-1	Sequence 1, Appli
27	c	176.8	17.7	59065	4	US-09-813-817-3	Sequence 3, Appli
28	c	176.8	17.7	59065	4	US-09-978-197-3	Sequence 3, Appli
29	c	176.6	17.6	16389	4	US-09-741-154-3	Sequence 3, Appli
30	c	176.6	17.6	98844	4	US-09-791-211-10	Sequence 10, Appl
31	c	176.6	17.6	246240	2	US-08-724-394A-20	Sequence 20, Appl
32	c	176.6	17.6	246240	2	US-08-724-394A-21	Sequence 21, Appl
33	c	176.6	17.6	246240	2	US-08-724-394A-22	Sequence 22, Appl
34		176.4	17.6	601	4	US-09-820-002-7	Sequence 7, Appli
35		176.4	17.6	601	4	US-09-820-002-8	Sequence 8, Appli
36		176.4	17.6	21784	4	US-09-820-002-3	Sequence 3, Appli
37	c	176.4	17.6	24707	4	US-09-740-027-3	Sequence 3, Appli
38		176.4	17.6	43069	4	US-09-292-542A-1	Sequence 1, Appli
39		176.4	17.6	80246	3	US-09-078-294-4	Sequence 4, Appli
40		176.4	17.6	80595	3	US-09-078-294-3	Sequence 3, Appli
41		176.4	17.5	36651	4	US-09-738-894A-3	Sequence 3, Appli
42		176.4	17.5	36651	4	US-09-964-469-3	Sequence 3, Appli
43	c	176.4	17.5	70000	4	US-09-851-896-3	Sequence 3, Appli
44		176.4	17.5	83450	4	US-09-811-469-3	Sequence 3, Appli
45	c	176.4	17.5	174493	4	US-09-804-471A-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-09-077-354B-3
; Sequence 3, Application US/09077354B
; Patent No. 6255096
;  GENERAL INFORMATION:
;    APPLICANT:  HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;
;    APPLICANT:  WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART
;    TITLE OF INVENTION:  SYNTHETIC MAMMALIAN
;    TITLE OF INVENTION:  N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES ENCODING SA
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCULLY, SCOTT, MURPHY & PRESSER
;      STREET:  400 GARDEN CITY PLAZA
;      CITY:  GARDEN CITY
;      STATE:  NEW YORK
;      COUNTRY:  UNITED STATES
;      ZIP:  11530
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/077,354B
;      FILING DATE:  22-APRIL-1999
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00747
;      FILING DATE:  22-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  POKALSKY, ANN R.
;      REGISTRATION NUMBER:  34,697
;      REFERENCE/DOCKET NUMBER:  12416
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  516 742 4343
;      TELEFAX:  516 742 4366
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10380 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    POSITION IN GENOME:
;      CHROMOSOME/SEGMENT:  Chromosome 17
;    FEATURE:
;      NAME/KEY:  exon 1
;      LOCATION:  990..1372
;    FEATURE:
```

```
;      NAME/KEY:  exon 2
;      LOCATION:  2115..2262
;    FEATURE:
;      NAME/KEY:  exon 3
;      LOCATION:  3056..3202
;    FEATURE:
;      NAME/KEY:  exon 4
;      LOCATION:  3387..3472
;    FEATURE:
;      NAME/KEY:  exon 5
;      LOCATION:  5667..5923
;    FEATURE:
;      NAME/KEY:  exon 6
;      LOCATION:  7745..8955
US-09-077-354B-3

  Query Match            19.0%;  Score 189.8;  DB 3;  Length 10380;
  Best Local Similarity   86.7%;  Pred. No. 1.5e-31;
  Matches  209;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64

Db       3540 TAAAAACTTAAGCTCTGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 3599

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124

Db       3600 AGGCCGAGTTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGT 3659

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184

Db       3660 GAAACCCTGTCTCTACTAAAAATATAAAAATTAGCTGGGCATGGTGGCACGCGCCTGTAA 3719

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244

Db       3720 TCCCATCTACTTAGGAGGCTGAGACAGGAGAATTGCTTAAACCTGGGAGGCAGACGTTGC 3779

Qy        245 A 245

Db       3780 A 3780

RESULT 2
US-09-791-211-10
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
;  APPLICANT:  Donna T. Ward
;  APPLICANT:  Andrew T. Watt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF WRN EXPRESSION
;  FILE REFERENCE:  RTS-0205
;  CURRENT APPLICATION NUMBER: US/09/791,211
;  CURRENT FILING DATE:  2001-02-23
;  NUMBER OF SEQ ID NOS:  90
; SEQ ID NO 10
;   LENGTH: 98844
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 24962
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 64383
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 65468
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 65469
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 65470
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 65471
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 87130
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 89049
;   OTHER INFORMATION: unknown
;   OTHER INFORMATION:
US-09-791-211-10

  Query Match            18.9%;  Score 188.8;  DB 4;  Length 98844;
  Best Local Similarity   86.7%;  Pred. No. 3.8e-31;
  Matches  206;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65

Db      82007 AAAAATAGCAAATTTTGGCCGGGCATGGTGGCTCACGCCTGTAGTCCCAGCACTTTGGGA 82066

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125

Db      82067 GGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGTCAACATGGTG 82126

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185

Db      82127 AAACCCTGTCTCTACTAAAAATACAACAATTAGCTGGGCGTGGTGGCAGGTGCCTGTAAT 82186

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      82187 CCCAGCTACTTGGGAGGCTGAGGCAGGGGAATTGCTTGAACTCGGGAGGTGGAGGTTGCA 82246

RESULT 3
US-09-800-960-3/c
; Sequence 3, Application US/09800960
; Patent No. 6387677
; GENERAL INFORMATION:
;  APPLICANT:  YE, Jane et al.
;  TITLE OF INVENTION:  ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE:  CL001158
;  CURRENT APPLICATION NUMBER: US/09/800,960
;  CURRENT FILING DATE:  2001-03-08
;  NUMBER OF SEQ ID NOS:  4
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 62804
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(62804)
;   OTHER INFORMATION: n = A,T,C or G
US-09-800-960-3

  Query Match            18.5%;  Score 185.6;  DB 4;  Length 62804;
  Best Local Similarity   89.3%;  Pred. No. 1.7e-30;
  Matches  200;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy         22 GGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGGA 81

Db      48950 GGGTGGGCACGGTGGCTCACACCTGTAATCTCAGCACTTTGGGAGGCCAAGGCGGGTGGA 48891

Qy         82 TCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTACT 141

Db      48890 TCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACT 48831

Qy        142 AAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGAG 201

Db      48830 AAAAATACAAAAATTAGCCGGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCGGGAG 48771
```

```
Qy        202 GCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db      48770 GCTGAGACAAGAGAATTGCTTGAACCTGGGAGGCAGAGATTGCA 48727

RESULT 4
US-08-579-445-26
; Sequence 26, Application US/08579445
; Patent No. 6566053
;  GENERAL INFORMATION:
;    APPLICANT:  Perucho, Manuel
;    APPLICANT:  Peinado, Miguel A.
;    APPLICANT:  Ionov, Yuri
;    APPLICANT:  Malkhosyan, Sergei
;    TITLE OF INVENTION:  Identification of Neoplasms by Detection
;    TITLE OF INVENTION:  of Genetic Deletions
;    NUMBER OF SEQUENCES:  27
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe, Martens, Olson & Bear
;      STREET:  620 Newport Center Drive, Sixteenth Floor
;      CITY:  Newport Beach
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/579,445
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/152,484
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kirkpatrick, Anita M.
;      REGISTRATION NUMBER:  32,617
;      REFERENCE/DOCKET NUMBER:  STRATAG.009A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  26:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  283 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-579-445-26

  Query Match             18.4%;  Score 184;  DB 4;  Length 283;
  Best Local Similarity   88.8%;  Pred. No. 1.3e-30;
  Matches  199;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         22 GGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGGA 81
Db          1 GGCTGGGCGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGA 60

Qy         82 TCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTACT 141
Db         61 TCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACT 120

Qy        142 AAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGAG 201
Db        121 AAAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGAG 180

Qy        202 GCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db        181 GCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGGAGGTTGCA 224

RESULT 5
US-08-133-629-8
; Sequence 8, Application US/08133629
; Patent No. 5597694
;  GENERAL INFORMATION:
;    APPLICANT:  Munroe, David J.
;    APPLICANT:  Housman, David E.
;    TITLE OF INVENTION:  AMPLIFICATION OF NUCLEIC ACIDS
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Wolf, Greenfield & Sacks, P.C.
;      STREET:  600 Atlantic Avenue
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  United States of America
;      ZIP:  02210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/133,629
;      FILING DATE:  07-OCT-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Greer, Helen
;      REGISTRATION NUMBER:  36,816
;      REFERENCE/DOCKET NUMBER:  M0828/7001
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-720-3500
;      TELEFAX:  617-720-2441
;      TELEX:  92-1742  EZEKIEL
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  282 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-133-629-8

  Query Match             18.3%;  Score 183.6;  DB 1;  Length 282;
  Best Local Similarity   88.4%;  Pred. No. 1.6e-30;
  Matches  198;  Conservative    1;  Mismatches   25;  Indels    0;  Gaps    0;

Qy         22 GGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGGA 81
Db          1 GGCTGGGCGTGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTGGA 60

Qy         82 TCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTACT 141
Db         61 TCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTACT 120

Qy        142 AAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGAG 201
Db        121 AAAAATACAAAAATTAGCCGGGCGTGGTGGCGCGCGCCTGTAATCCCAGCTACTCGGGAG 180

Qy        202 GCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db        181 GCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGTGGAGGYTGCA 224

RESULT 6
US-09-539-333D-37
; Sequence 37, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
```

```
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 37
;   LENGTH: 1154
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: 1..719
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 720..1118
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: 1119..1154
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 1131..1136
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 191
;   OTHER INFORMATION: 8-121-197  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 313
;   OTHER INFORMATION: 8-122-271  : deletion of CAAA
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 314
;   OTHER INFORMATION: 8-122-272  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 368
;   OTHER INFORMATION: 8-122-326  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 390
;   OTHER INFORMATION: 8-123-55  : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 814
;   OTHER INFORMATION: 8-128-61  : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 821
;   OTHER INFORMATION: 8-128-68  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 822
;   OTHER INFORMATION: 8-128-69  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 838
;   OTHER INFORMATION: 8-128-85  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 897
;   OTHER INFORMATION: 8-129-50  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 908
;   OTHER INFORMATION: 8-129-60  : deletion of A
US-09-539-333D-37

  Query Match             18.3%;  Score 183.6;  DB 4;  Length 1154;
  Best Local Similarity   85.7%;  Pred. No. 2.1e-30;
  Matches  204;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          8 AAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGG 67

Db        503 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG 562

Qy         68 CCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127

Db        563 CCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGGTGAA 622

Qy        128 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCC 187

Db        623 ACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGTGCCTGTAGTCC 682

Qy        188 CAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db        683 CAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCA 740


RESULT 7
US-09-539-333D-36
; Sequence 36, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 36
;   LENGTH: 1301
```

```
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: 1..899
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 900..1265
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: 1266..1301
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 1277..1282
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 191
;   OTHER INFORMATION: 8-121-187  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 313
;   OTHER INFORMATION: 8-122-271  : deletion of CAAA
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 314
;   OTHER INFORMATION: 8-122-272  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 368
;   OTHER INFORMATION: 8-122-326  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 390
;   OTHER INFORMATION: 8-123-55  : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 806
;   OTHER INFORMATION: 8-127-28  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 897
;   OTHER INFORMATION: 8-127-119  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 937
;   OTHER INFORMATION: 8-127-159  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 961
;   OTHER INFORMATION: 8-128-61  : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 968
;   OTHER INFORMATION: 8-128-68  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 969
;   OTHER INFORMATION: 8-128-69  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 985
;   OTHER INFORMATION: 8-128-85  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1044
;   OTHER INFORMATION: 8-129-50  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1055
;   OTHER INFORMATION: 8-129-60  : deletion of A
US-09-539-333D-36

  Query Match             18.3%;  Score 183.6;  DB 4;  Length 1301;
  Best Local Similarity   85.7%;  Pred. No. 2.1e-30;
  Matches  204;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          8 AAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGG 67
Db        503 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG 562

Qy         68 CCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127
Db        563 CCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGGTGAA 622

Qy        128 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCC 187
Db        623 ACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGTGCCTGTAGTCC 682

Qy        188 CAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db        683 CAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCA 740


RESULT 8
US-09-539-333D-40
; Sequence 40, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 40
;   LENGTH: 1386
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: 1..984
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 985..1350
;   FEATURE:
;   NAME/KEY: 3'UTR
;   LOCATION: 1351..1386
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 1362..1367
;   FEATURE:
;   NAME/KEY: allele
```

```
;   LOCATION: 191
;   OTHER INFORMATION: 8-121-187  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 398
;   OTHER INFORMATION: 8-122-271  : deletion of CAAA
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 399
;   OTHER INFORMATION: 8-122-272  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 453
;   OTHER INFORMATION: 8-122-326  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 475
;   OTHER INFORMATION: 8-123-55  : polymorphic base A or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 891
;   OTHER INFORMATION: 8-127-28  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 982
;   OTHER INFORMATION: 8-127-119  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1022
;   OTHER INFORMATION: 8-127-159  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1046
;   OTHER INFORMATION: 8-128-61  : polymorphic base G or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1053
;   OTHER INFORMATION: 8-128-68  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1054
;   OTHER INFORMATION: 8-128-69  : polymorphic base A or G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1070
;   OTHER INFORMATION: 8-128-85  : polymorphic base A or C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1129
;   OTHER INFORMATION: 8-129-50  : polymorphic base C or T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 1140
;   OTHER INFORMATION: 8-129-60  : deletion of A
US-09-539-333D-40

  Query Match             18.3%;  Score 183.6;  DB 4,  Length 1386;
  Best Local Similarity   85.7%;  Pred. No. 2.1e-30;
  Matches  204;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          8 AAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGG 67
              ||  |  |    |  | |||  |||  .|||:::|||  ||.::||.::||  |||:|||  |
Db        588 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG 647

Qy         68 CCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127
              || ||: :||| :|||||:|:|||: |||:|:||||  |||: |||: ||    |||: |:.
Db        648 CCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGGTGAA 707

Qy        128 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCC 187
              :|||  ::|||||:|:||||:|:|||:||||:  ||: |||   ||: |||: |||| |||
Db        708 ACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGTGCCTGTAGTCC 767

Qy        188 CAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
              :||:|||  :|||:|:||: ||| :|||| : ||||.:  |||:  .||||  : ||
Db        768 CAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCA 825

RESULT 9
US-09-679-409-1
; Sequence 1, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
;  APPLICANT:  Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION:  SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE:  53.US15.CIP
;  CURRENT APPLICATION NUMBER: US/09/679,409
;  CURRENT FILING DATE:  2000-10-03
;  PRIOR APPLICATION NUMBER:  09/539,333
;  PRIOR FILING DATE:  2000-03-03
;  PRIOR APPLICATION NUMBER:  09/416,384
;  PRIOR FILING DATE:  1999-10-12
;  PRIOR APPLICATION NUMBER:  60/168,098
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  134
;  SOFTWARE:  Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 199122..201122
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 201123..201234
;   OTHER INFORMATION: exon S
;   NAME/KEY: exon
;   LOCATION: 201123..201560
;   OTHER INFORMATION: exon S2
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U
;   NAME/KEY: exon
;   LOCATION: 216836..216994
;   OTHER INFORMATION: exon V
;   NAME/KEY: exon
;   LOCATION: 216836..217077
;   OTHER INFORMATION: exon V2
;   NAME/KEY: exon
;   LOCATION: 217671..217764
;   OTHER INFORMATION: exon V1
;   NAME/KEY: exon
;   LOCATION: 227655..227736
;   OTHER INFORMATION: exon V4
;   NAME/KEY: exon
;   LOCATION: 238715..238919
;   OTHER INFORMATION: exon V3
;   NAME/KEY: exon
;   LOCATION: 240440..240673
;   OTHER INFORMATION: exon W
;   NAME/KEY: exon
;   LOCATION: 240440..241153
;   OTHER INFORMATION: exon W2
;   NAME/KEY: exon
;   LOCATION: 241072..241291
;   OTHER INFORMATION: exon X
;   NAME/KEY: exon
;   LOCATION: 244353..244561
;   OTHER INFORMATION: exon Y
```

```
;    NAME/KEY: exon
;    LOCATION: 246273..247802
;    OTHER INFORMATION: exon Z
;    NAME/KEY: misc_feature
;    LOCATION: 247803..249803
;    OTHER INFORMATION: 3'regulatory region
;    NAME/KEY: allele
;    LOCATION: 8316
;    OTHER INFORMATION: 99-27943-150  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 21672
;    OTHER INFORMATION: 99-27935-193  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 65485
;    OTHER INFORMATION: 8-128-33  : polymorphic base C or T
;    NAME/KEY: allele
;    LOCATION: 95396
;    OTHER INFORMATION: 99-31960-363  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 107281
;    OTHER INFORMATION: 99-24656-260  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 160640
;    OTHER INFORMATION: 99-24639-163  : polymorphic base A or C
;    NAME/KEY: allele
;    LOCATION: 160876
;    OTHER INFORMATION: 99-24634-108  : polymorphic base A or T
;    NAME/KEY: allele
;    LOCATION: 168974
;    OTHER INFORMATION: 99-7652-162  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 170810
;    OTHER INFORMATION: 99-16100-147  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 173358
;    OTHER INFORMATION: 99-5862-167  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 189957
;    OTHER INFORMATION: 99-5919-215  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 197163
;    OTHER INFORMATION: 99-24658-410  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 200778
;    OTHER INFORMATION: 8-303-235  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 202651
;    OTHER INFORMATION: 8-300-221  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 202679
;    OTHER INFORMATION: 8-300-193  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 203378
;    OTHER INFORMATION: 8-299-128  : polymorphic base A or T
;    NAME/KEY: allele
;    LOCATION: 204138
;    OTHER INFORMATION: 8-296-213  : polymorphic base A or T
;    NAME/KEY: allele
;    LOCATION: 204605
;    OTHER INFORMATION: 8-252-190  : polymorphic base C or T
;    NAME/KEY: allele
;    LOCATION: 204934
;    OTHER INFORMATION: 99-24644-194  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 205206
;    OTHER INFORMATION: 8-295-248  : polymorphic base A or C
;    NAME/KEY: allele
;    LOCATION: 205329
;    OTHER INFORMATION: 8-295-125  : polymorphic base C or T
;    NAME/KEY: allele
;    LOCATION: 206064
;    OTHER INFORMATION: 8-293-130  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 206545
;    OTHER INFORMATION: 8-292-198  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 207313
;    OTHER INFORMATION: 8-251-322  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 208265
;    OTHER INFORMATION: 8-289-322  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 208960
;    OTHER INFORMATION: 8-287-249  : polymorphic base C or T
;    NAME/KEY: allele
;    LOCATION: 209123
;    OTHER INFORMATION: 8-287-86  : polymorphic base A or T
;    NAME/KEY: allele
;    LOCATION: 209631
;    OTHER INFORMATION: 8-285-319  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 210361
;    OTHER INFORMATION: 8-283-278  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 210463
;    OTHER INFORMATION: 8-283-176  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 210486
;    OTHER INFORMATION: 8-283-153  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 210583
;    OTHER INFORMATION: 8-283-56  : polymorphic base C or T
;    NAME/KEY: allele
;    LOCATION: 210879
;    OTHER INFORMATION: 8-282-345  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 210964
;    OTHER INFORMATION: 8-282-260  : polymorphic base G or T
;    NAME/KEY: allele
;    LOCATION: 210979
;    OTHER INFORMATION: 8-282-245  : polymorphic base A or C
;    NAME/KEY: allele
;    LOCATION: 211050
;    OTHER INFORMATION: 8-282-174  : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
;    NAME/KEY: allele
;    LOCATION: 211132
;    OTHER INFORMATION: 8-282-92  : polymorphic base A or T
;    NAME/KEY: allele
;    LOCATION: 211247
;    OTHER INFORMATION: 8-281-367  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 211315
;    OTHER INFORMATION: 8-281-299  : polymorphic base A or G
;    NAME/KEY: allele
;    LOCATION: 211366
;    OTHER INFORMATION: 8-281-248  : polymorphic base G or C
;    NAME/KEY: allele
;    LOCATION: 212520
;    OTHER INFORMATION: 8-279-197  : polymorphic base A or C
;    NAME/KEY: allele
;    LOCATION: 212821
;    OTHER INFORMATION: 8-278-289  : polymorphic base C or T
;    NAME/KEY: allele

  Query Match             18.3%;  Score 183.6;  DB 4;  Length 319608;
  Best Local Similarity   85.7%;  Pred. No. 6e-30;
  Matches  204;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          8 AAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGG 67
              ||  | |   | | |||  |||  ||||||||||  ||||||||||||||||||||||||
Db      29980 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG 30039

Qy         68 CCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127
              || ||| |||| ||||||||||||| ||||||||| ||||| |||| |||  |||| |||
Db      30040 CCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGGTGAA 30099
```

```
Qy        128 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCC 187
              ||||  |||||| |||| |||||||||||||||    |||   ||   ||.||  | | |
Db      30100 ACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGTGCCTGTAGTCC 30159

Qy        188 CAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
              |||||.| |||| ||||  ||| | |||  ||| .|||  |||   ||.||| .||
Db      30160 CAGCTACTCGGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCA 30217


RESULT 10
US-09-539-333D-1
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
;  APPLICANT: Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE: GENSET.047AUS
;  CURRENT APPLICATION NUMBER: US/09/539,333D
;  CURRENT FILING DATE:  2000-03-30
;  PRIOR APPLICATION NUMBER: US 60/126,903
;  PRIOR FILING DATE: 1999-03-30
;  PRIOR APPLICATION NUMBER: US 60/131,971
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/132,065
;  PRIOR FILING DATE: 1999-04-30
;  PRIOR APPLICATION NUMBER: US 60/143,928
;  PRIOR FILING DATE: 1999-07-14
;  PRIOR APPLICATION NUMBER: US 60/145,915
;  PRIOR FILING DATE: 1999-07-27
;  PRIOR APPLICATION NUMBER: US 60/146,453
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/146,452
;  PRIOR FILING DATE: 1999-07-29
;  PRIOR APPLICATION NUMBER: US 60/162,288
;  PRIOR FILING DATE: 1999-10-28
;  PRIOR APPLICATION NUMBER: US 09/416,384
;  PRIOR FILING DATE: 1999-10-12
;  NUMBER OF SEQ ID NOS: 231
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31..1107
;   OTHER INFORMATION: 5'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1108..1285
;   OTHER INFORMATION: exon A g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14877..14920
;   OTHER INFORMATION: exon B g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18778..18862
;   OTHER INFORMATION: exon Bbis g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25593..25740
;   OTHER INFORMATION: exon C g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29388..29502
;   OTHER INFORMATION: exon D g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29367..30282
;   OTHER INFORMATION: exon E g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 64666..64812
;   OTHER INFORMATION: exon F g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 65505..65853
;   OTHER INFORMATION: exon G g35018 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 65854..67854
;   OTHER INFORMATION: 3'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 94124..94964
;   OTHER INFORMATION: exon g35017
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 201188..201234
;   OTHER INFORMATION: exon S g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216836..216915
;   OTHER INFORMATION: exon V g35030 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 213818..215818
;   OTHER INFORMATION: 3'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215941
;   OTHER INFORMATION: exon R complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215975
;   OTHER INFORMATION: exon Rbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..216952
;   OTHER INFORMATION: exon Qbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..217061
;   OTHER INFORMATION: exon Q complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 217027..217061
;   OTHER INFORMATION: exon Q1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 229647..229742
;   OTHER INFORMATION: exon X complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 230408..230721
;   OTHER INFORMATION: exon P complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231272..231412
;   OTHER INFORMATION: exon Obis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
```

```
;   LOCATION: 231787..231880
;   OTHER INFORMATION: exon O2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231870..231879
;   OTHER INFORMATION: exon O1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 234174..234321
;   OTHER INFORMATION: exon C complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 237406..237428
;   OTHER INFORMATION: exon Nbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239807
;   OTHER INFORMATION: exon N2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239853
;   OTHER INFORMATION: exon N complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240569
;   OTHER INFORMATION: exon M1117 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240596
;   OTHER INFORMATION: exon M1090 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240617
;   OTHER INFORMATION: exon M1069 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240644
;   OTHER INFORMATION: exon MS2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240824
;   OTHER INFORMATION: exon M862 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240994
;   OTHER INFORMATION: exon M692 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..241685
;   OTHER INFORMATION: exon M1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240800..240993
;   OTHER INFORMATION: exon MS1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 241686..243685
;   OTHER INFORMATION: 5'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 290652..292652
;   OTHER INFORMATION: 3'regulatory region g34665 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 292653..292841

  Query Match             18.3%;  Score 193.2;  DB 4;  Length 319608;
  Best Local Similarity   85.3%;  Pred. No. 7.3e-30;
  Matches  203;  Conservative    1;  Mismatches   34;  Indels    0;  Gaps    0;

Qy          8 AAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGG 67
Db      29980 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGG 30039

Qy         68 CCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127
Db      30040 CCGAGGCGGGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAATATGGTGAA 30099

Qy        128 ACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCC 187
Db      30100 ACCCCATCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCAGGTGCCTGTAGTCC 30159

Qy        188 CAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db      30160 CAGCTACTCRGGAGGCTGAGACAGGAGAATTGCTTGAACCTGGGAGGCGGAGGTTGCA 30217


RESULT 11
US-09-750-580-1/c
; Sequence 1, Application US/09750580
; Patent No. 6455280
; GENERAL INFORMATION:
;  APPLICANT:  Yen, Frances
;  APPLICANT:  Denison, Blake
;  APPLICANT:  Bour, Barbara
;  APPLICANT:  Bihain, Bernard
;  APPLICANT:  Dumas Milne Edwards, Jean-Baptiste
;  APPLICANT:  Duclert, Aymeric
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Ebbets-Reed, Dana
;  APPLICANT:  Salter-Cid, Luisa
;  TITLE OF INVENTION:  METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
;  FILE REFERENCE:  89.US2.CIP
;  CURRENT APPLICATION NUMBER: US/09/750,580
;  CURRENT FILING DATE:  2000-12-28
;  PRIOR APPLICATION NUMBER:  US 09/599,362
;  PRIOR FILING DATE:  2000-06-21
;  PRIOR APPLICATION NUMBER:  PCT/IB00/0101
;  PRIOR FILING DATE:  2000-06-21
;  PRIOR APPLICATION NUMBER:  PCT/IB99/02058
;  PRIOR FILING DATE:  1999-12-20
;  PRIOR APPLICATION NUMBER:  US 49/469/099
;  PRIOR FILING DATE:  1999-12-21
;  PRIOR APPLICATION NUMBER:  US 60/113,686
;  PRIOR FILING DATE:  1998-12-22
;  PRIOR APPLICATION NUMBER:  US 60/141,032
;  PRIOR FILING DATE:  1999-06-25
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  Patent.pm
; SEQ ID NO 1
;   LENGTH: 81001
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 10946..12946
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 12947..12958
;   OTHER INFORMATION: exon 1
;   NAME/KEY: exon
;   LOCATION: 13470..13526
;   OTHER INFORMATION: exon 2
;   NAME/KEY: exon
;   LOCATION: 13641..13752
;   OTHER INFORMATION: exon 3
;   NAME/KEY: exon
;   LOCATION: 14271..15968
;   OTHER INFORMATION: exon 4
;   NAME/KEY: misc_feature
;   LOCATION: 15969..17969
;   OTHER INFORMATION: 3'regulatory region
;   NAME/KEY: allele
;   LOCATION: 1239
;   OTHER INFORMATION: 20-828-311  : polymorphic base C or T
;   NAME/KEY: allele
```

```
;   LOCATION: 12347
;   OTHER INFORMATION: 17-42-319   : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 15241
;   OTHER INFORMATION: 17-41-250   : polymorphic base C or T
;   NAME/KEY: allele
;   LOCATION: 42218
;   OTHER INFORMATION: 20-841-149  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 45442
;   OTHER INFORMATION: 20-842-115  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 77058
;   OTHER INFORMATION: 20-853-415  : polymorphic base C or T
;   NAME/KEY: primer_bind
;   LOCATION: 929..949
;   OTHER INFORMATION: 20-828.pu
;   NAME/KEY: primer_bind
;   LOCATION: 1357..1377
;   OTHER INFORMATION: 20-828.rp  complement
;   NAME/KEY: primer_bind
;   LOCATION: 12029..12050
;   OTHER INFORMATION: 17-42.pu
;   NAME/KEY: primer_bind
;   LOCATION: 12581..12603
;   OTHER INFORMATION: 17-42.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 14992..15012
;   OTHER INFORMATION: 17-41.pu
;   NAME/KEY: primer_bind
;   LOCATION: 15460..15482
;   OTHER INFORMATION: 17-41.rp complement
;   NAME/KEY: primer_bind
;   LOCATION: 42070..42090
;   OTHER INFORMATION: 20-841.pu
;   NAME/KEY: primer_bind
;   LOCATION: 42572..42591
;   OTHER INFORMATION: 20-841.rp  complement
;   NAME/KEY: primer_bind
;   LOCATION: 45328..45347
;   OTHER INFORMATION: 20-842.pu
;   NAME/KEY: primer_bind
;   LOCATION: 45863..45883
;   OTHER INFORMATION: 20-842.rp  complement
;   NAME/KEY: primer_bind
;   LOCATION: 76644..76664
;   OTHER INFORMATION: 20-853.pu
;   NAME/KEY: primer_bind
;   LOCATION: 77166..77185
;   OTHER INFORMATION: 20-853.rp  complement
;   NAME/KEY: primer_bind
;   LOCATION: 1220..1238
;   OTHER INFORMATION: 20-828-311.mis
;   NAME/KEY: primer_bind
;   LOCATION: 1240..1258
;   OTHER INFORMATION: 20-828-311.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 12328..12346
;   OTHER INFORMATION: 17-42-319.mis
;   NAME/KEY: primer_bind
;   LOCATION: 12348..12366
;   OTHER INFORMATION: 17-42-319.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 15222..15240
;   OTHER INFORMATION: 17-41-250.mis
;   NAME/KEY: primer_bind
;   LOCATION: 15242..15260
;   OTHER INFORMATION: 17-41-250.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 42199..42217
;   OTHER INFORMATION: 20-841-149.mis
;   NAME/KEY: primer_bind
;   LOCATION: 42219..42237
;   OTHER INFORMATION: 20-841-149.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 45423..45441
;   OTHER INFORMATION: 20-842-115.mis
;   NAME/KEY: primer_bind
;   LOCATION: 45443..45461
;   OTHER INFORMATION: 20-842-115.mis complement
;   NAME/KEY: primer_bind
;   LOCATION: 77039..77057
;   OTHER INFORMATION: 20-853-415.mis
;   NAME/KEY: primer_bind
;   LOCATION: 77059..77077
;   OTHER INFORMATION: 20-853-415.mis complement
;   NAME/KEY: misc_binding
;   LOCATION: 1227..1251
;   OTHER INFORMATION: 20-828-311.probe
;   NAME/KEY: misc_binding
;   LOCATION: 12335..12359
;   OTHER INFORMATION: 17-42-319.probe
;   NAME/KEY: misc_binding
;   LOCATION: 15229..15253
;   OTHER INFORMATION: 17-41-250.probe
;   NAME/KEY: misc_binding
;   LOCATION: 42206..42230
;   OTHER INFORMATION: 20-841-149.probe
;   NAME/KEY: misc_binding
;   LOCATION: 45430..45454
;   OTHER INFORMATION: 20-842-115.probe
;   NAME/KEY: misc_binding
;   LOCATION: 77046..77070
;   OTHER INFORMATION: 20-853-415.probe
US-09-750-580-1

  Query Match             18.2%;  Score 182.4;  DB 4;  Length 81001;
  Best Local Similarity   85.0%;  Pred. No. 8.3e-30;
  Matches  204;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65

Db      65052 AATAAGGAAATTGTGTGGCCAGGTGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 64993

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125

Db      64992 GGCCAAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 64933

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185

Db      64932 AAACCCCGTCTCTATTAAAAATACAAAAATTAGCCGGGCATGGTGGCAGGAGCCTGTAAT 64873

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      64872 CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCACTTGAACCCGGGAGGCGGAGGTTGCA 64813


RESULT 12
US-09-499-522-1
; Sequence 1, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT: Bihain, Bernard
;  TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
;  FILE REFERENCE: GENSET.053AUS
;  CURRENT APPLICATION NUMBER: US/09/499,522
;  CURRENT FILING DATE: 2000-02-10
;  EARLIER APPLICATION NUMBER: US 60/119,592
;  EARLIER FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: US 60/144,784
;  EARLIER FILING DATE: 1999-07-20
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
```

```
;   LENGTH: 23187
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 2001..2356
;   OTHER INFORMATION: exon1
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 3540..3884
;   OTHER INFORMATION: exon2
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 12163..12282
;   OTHER INFORMATION: exon3
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 15144..15200
;   OTHER INFORMATION: exon4
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 15765..15911
;   OTHER INFORMATION: exon5
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19579..19752
;   OTHER INFORMATION: exon6
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 19899..19958
;   OTHER INFORMATION: exon7
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 20056..20187
;   OTHER INFORMATION: exon8
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 20329..20957
;   OTHER INFORMATION: exon9
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 21047..21187
;   OTHER INFORMATION: exon10
;   FEATURE:
;   NAME/KEY: polyA_signal
;   LOCATION: 21168..21173
;   OTHER INFORMATION: AATAAA
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..2000
;   OTHER INFORMATION: potential 5'regulatory region
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 22324..23187
;   OTHER INFORMATION: homology with USF2 gene in ref: embl Y07661
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 523..544
;   OTHER INFORMATION: upstream amplification primer 17-2
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1047..1068
;   OTHER INFORMATION: downstream amplification primer 17-2 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 946..963
;   OTHER INFORMATION: upstream amplification primer 99-4576
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1385..1402
;   OTHER INFORMATION: downstream amplification primer 99-4576 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1096..1115
;   OTHER INFORMATION: upstream amplification primer 9-19
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1616..1635
;   OTHER INFORMATION: downstream amplification primer 9-19 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 1602..1621
;   OTHER INFORMATION: upstream amplification primer 9-20
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2074..2093
;   OTHER INFORMATION: downstream amplification primer 9-20 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2036..2053
;   OTHER INFORMATION: upstream amplification primer 99-4557
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2563..2580
;   OTHER INFORMATION: downstream amplification primer 99-4557 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2084..2102
;   OTHER INFORMATION: upstream amplification primer 9-1
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2483..2500
;   OTHER INFORMATION: downstream amplification primer 9-1 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2470..2489
;   OTHER INFORMATION: upstream amplification primer 9-21 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 2062..2081
;   OTHER INFORMATION: downstream amplification primer 9-21
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3455..3474
;   OTHER INFORMATION: upstream amplification primer 9-3
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3882..3901
;   OTHER INFORMATION: downstream amplification primer 9-3 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 3775..3792
;   OTHER INFORMATION: upstream amplification primer 99-4558
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4336..4356
;   OTHER INFORMATION: downstream amplification primer 99-4558 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4902..4920
;   OTHER INFORMATION: upstream amplification primer 99-14419 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 4444..4463
;   OTHER INFORMATION: downstream amplification primer 99-14419
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 6638..6655
;   OTHER INFORMATION: upstream amplification primer 99-4577
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 7072..7089
;   OTHER INFORMATION: downstream amplification primer 99-4577 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 7995..8012
```

```
;   OTHER INFORMATION: upstream amplification primer 99-4559
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 8576..8593
;   OTHER INFORMATION: downstream amplification primer 99-4559 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 9622..9639
;   OTHER INFORMATION: upstream amplification primer 99-3148
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10023..10040
;   OTHER INFORMATION: downstream amplification primer 99-3148 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 9964..9981
;   OTHER INFORMATION: upstream amplification primer 99-4560
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10546..10563
;   OTHER INFORMATION: downstream amplification primer 99-4560 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10996..11015
;   OTHER INFORMATION: upstream amplification primer 99-14411 , complement
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 10492..10512
;   OTHER INFORMATION: downstream amplification primer 99-14411
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 11972..11990
;   OTHER INFORMATION: upstream amplification primer 99-4561
;   FEATURE:
;   NAME/KEY: primer_bind
;   LOCATION: 12481..12501
;   OTHER INFORMATION: downstream amplification primer 99-4561 , complement

  Query Match             18.2%;  Score 181.8;  DB 4;  Length 23187;
  Best Local Similarity   88.0%;  Pred. No. 8.8e-30;
  Matches  198;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy         21 AGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGG 80

Db       5591 AGTCTGGGCACGGTGGCTCACGTCTGTAATCTCAGCACTTTGGGAGGCCAAGGTGGGTGG 5650

Qy         81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTAC 140

Db       5651 ATCACCTGAGGTCAGGAGTTCCAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCTAC 5710

Qy        141 TAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGA 200

Db       5711 TAAAAATACAAAGATTAGCCAGGCATGGTGGCACATGCCTGTAATCCCAGCTACTCAGGA 5770

Qy        201 GGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db       5771 GGCTGAGGCATGAGAATCGCTTGAACCCAGGAGGCAGAGGTTGCA 5815

RESULT 13
US-09-820-002-3/c
; Sequence 3, Application US/09820002
; Patent No. 6482630
; GENERAL INFORMATION:
;  APPLICANT: Gan, Weiniu
;  APPLICANT:  Ye, Jane
;  APPLICANT:  DiFrancesco, Valentina
;  APPLICANT:  Beasley, Ellen
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001194
;  CURRENT APPLICATION NUMBER: US/09/820,002
;  CURRENT FILING DATE:  2001-03-29
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 21784
;   TYPE: DNA
;   ORGANISM: HUMAN
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(21784)
;   OTHER INFORMATION: n = A,T,C or G
US-09-820-002-3

  Query Match             18.0%;  Score 180.6;  DB 4;  Length 21784;
  Best Local Similarity   84.0%;  Pred. No. 1.6e-29;
  Matches  204;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy          3 TATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTG 62

Db      18354 TAAAAAAAAAACCACAATAGGCTGGGCACAGTGGCTCATGCCTGTAATACCAGCACTTTG 18295

Qy         63 GGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATG 122

Db      18294 GGAGGCCGAGGCAGGAGGATCACTTGATGTCAGGAGTTTGAGACCAGCCTGGCCAACATG 18235

Qy        123 GCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGT 182

Db      18234 GCGAAACCTCATCTCTACTAAAAATACAAAAATTAGCCGGGCATGGTGGCACGCACCTGT 18175

Qy        183 AATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTT 242

Db      18174 AATCCCAGCTACTCGGGAGACTGAGGCAGAAGAATCACTTGAACCCAGGAGGTGGAGGTT 18115

Qy        243 GCA 245

Db      18114 GCA 18112

RESULT 14
US-09-443-184-35
; Sequence 35, Application US/09443184A
; Patent No. 6372431
; GENERAL INFORMATION:
;  APPLICANT: Cunningham, Mary Jane
;  APPLICANT: Zweiger, Gary
;  APPLICANT: Kaser, Matthew R.
;  APPLICANT: Panzer, Scott
;  APPLICANT: Seilhammer, Jeffrey J.
;  APPLICANT: Yue, Henry
;  APPLICANT: Baughn, Mariah
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lal, Preeti
;  TITLE OF INVENTION: MAMMALIAN TOXICOLOGICAL RESPONSE MARKERS
;  FILE REFERENCE: PC-0007 US
;  CURRENT APPLICATION NUMBER: US/09/443,184A
;  CURRENT FILING DATE: 1999-11-19
;  NUMBER OF SEQ ID NOS: 136
;  SOFTWARE: PERL Program
; SEQ ID NO 35
;   LENGTH: 1762
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6372431 2742442CB1
US-09-443-184-35

  Query Match             18.0%;  Score 180.2;  DB 4;  Length 1762;
  Best Local Similarity   84.2%;  Pred. No. 1.2e-29;
  Matches  203;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64
```

```
Db       1458 TTAAAGTAAAAACTTTGGGCTGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTGGGG 1517

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124

Db       1518 AAGCTGAGGTGGGCAGATCAGTTGAGGTCAGGAGTTCTAGACCAGCCTGGTCAACATGGC 1577

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184

Db       1578 GAAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCACACACCTGTAA 1637

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244

Db       1638 TCCCAGTTACTTGGGAGGCTGAGGCACAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGC 1697

Qy        245 A 245

Db       1698 A 1698
```

RESULT 15
US-09-491-356C-1/c
; Sequence 1, Application US/09491356C
; Patent No. 6566061
; GENERAL INFORMATION:
; APPLICANT: Philibert, Robert A.
; APPLICANT: Ginns, Edward I.
; APPLICANT: Delisi, Lynn
; TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ13
; FILE REFERENCE: 9465.6USC1
; CURRENT APPLICATION NUMBER: US/09/491,356C
; CURRENT FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: PCT/US99/09365
; PRIOR FILING DATE: 1999-04-29
; PRIOR APPLICATION NUMBER: 60/083,465
; PRIOR FILING DATE: 1998-04-29
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 55298
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (485)..(485)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (838)..(838)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (16728)..(16728)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (22750)..(22750)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (22756)..(22756)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (28519)..(28519)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (44804)..(44804)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (45002)..(45002)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (54049)..(54049)
; OTHER INFORMATION: n is not determined
; NAME/KEY: misc_feature
; LOCATION: (54226)..(54226)
; OTHER INFORMATION: n is not determined
US-09-491-356C-1

```
  Query Match           18.0%;  Score 180;  DB 4;  Length 55298;
  Best Local Similarity  83.6%;  Pred. No. 2.5e-29;
  Matches  204;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          2 ATATAAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTT 61

Db      43198 ATCTAGAAAAGCACCTTAGAGGCTGGGCGTGATGGCTCATGCCTGTAATCCCAGCACTTT 43139

Qy         62 GGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCAT 121

Db      43138 GGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACAT 43079

Qy        122 GGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTG 181

Db      43078 GGTGAAACCCCGTCTCTACTAAAAATACAAAAATATGCCGGGTGTGGTGGCAGGCGCCTG 43019

Qy        182 TAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGT 241

Db      43018 TAATCCCAGCTACTTGGGAGGCTGAGGCAGGAGAATTGCTTTAACCGGGGAGGCGGAGGT 42959

Qy        242 TGCA 245

Db      42958 TGTA 42955
```

Search completed: October 24, 2003, 15:43:01
Job time : 66.6458 secs

GenCore version 5.1.6

CM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 246.585 Seconds
(without alignments)
10886.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....acctgattctggtcttctttt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	191.4	19.1	3143	12	US-10-275-910-1	Sequence 1, Appli
2	191.4	19.1	30573	11	US-09-764-891-6869	Sequence 6869, Ap
3	191.4	19.1	32195	11	US-09-764-691-6868	Sequence 6868, Ap
4	190.8	19.1	7677	12	US-09-814-353-21993	Sequence 21993, A
5	189.8	19.0	10380	11	US-09-836-613-3	Sequence 3, Appli
6	189.4	18.9	1325	13	US-10-027-632-123771	Sequence 123771,
7	189.4	18.9	1325	13	US-10-027-632-123772	Sequence 123772,
8	189.4	18.9	1325	13	US-10-027-632-123773	Sequence 123773,
c 9	189.4	18.9	2490	13	US-10-027-632-102144	Sequence 102144,
10	189.4	18.9	2490	13	US-10-027-632-111607	Sequence 111607,
11	188.8	18.9	614	13	US-10-027-632-306904	Sequence 306904,
12	188.8	18.9	614	13	US-10-027-632-306905	Sequence 306905,
13	188.8	18.9	944	13	US-10-027-632-77917	Sequence 77917, A
14	188.8	18.9	17792	9	US-09-764-869-1599	Sequence 1599, Ap
15	188.8	18.9	17792	14	US-10-091-504-1599	Sequence 1599, Ap
c 16	188.6	18.8	19616	10	US-09-764-877-3220	Sequence 3220, Ap
c 17	187.6	18.7	11754	11	US-09-984-827-5	Sequence 5, Appli
18	186.6	18.6	17335	10	US-09-764-847-1280	Sequence 1280, Ap
19	186.6	18.6	17335	14	US-10-092-154-1280	Sequence 1280, Ap
20	186.6	18.6	19882	10	US-09-764-847-1281	Sequence 1281, Ap
21	186.6	18.6	19882	14	US-10-092-154-1281	Sequence 1281, Ap
c 22	186.6	18.6	116840	14	US-10-020-141-3	Sequence 3, Appli
23	186.6	18.6	2691139	14	US-10-067-514-1	Sequence 1, Appli
c 24	186.4	18.6	13467	10	US-09-764-868-1330	Sequence 1330, Ap
c 25	186.4	18.6	13467	10	US-09-764-868-1498	Sequence 1498, Ap
c 26	186.4	18.6	13467	11	US-09-764-891-8632	Sequence 8632, Ap
27	185.8	18.6	790	13	US-10-027-632-145713	Sequence 145713,
c 28	185.6	18.6	32190	9	US-09-764-869-2209	Sequence 2209, Ap
c 29	185.8	18.6	32190	14	US-10-091-504-2209	Sequence 2209, Ap
c 30	185.6	18.5	13058	10	US-09-764-846-303	Sequence 303, App
c 31	185.6	18.5	13058	14	US-10-091-483-303	Sequence 303, App
32	185.6	18.5	25001	12	US-10-017-161-2063	Sequence 2063, Ap
c 33	185.6	18.5	62804	13	US-10-096-960-3	Sequence 3, Appli
c 34	185.2	18.5	32246	9	US-09-764-860-802	Sequence 802, Ap
c 35	185.2	18.5	32246	14	US-10-074-095-802	Sequence 802, App
c 36	185.2	18.5	41936	10	US-09-967-768A-116	Sequence 116, App
c 37	185.2	18.5	162025	12	US-10-272-665-35	Sequence 35, Appl
c 38	185.2	18.5	162025	12	US-10-272-665-36	Sequence 36, Appl
c 39	185.2	18.5	162025	12	US-10-273-321-35	Sequence 35, Appl
c 40	185.2	18.5	162025	12	US-10-273-321-36	Sequence 36, Appl
c 41	185.2	18.5	162025	12	US-10-272-756-35	Sequence 35, Appl
c 42	185.2	18.5	162025	12	US-10-272-756-36	Sequence 36, Appl
43	185	18.5	421	11	US-09-918-995-4270	Sequence 4270, Ap
c 44	185	18.5	637	13	US-10-027-632-218180	Sequence 218180,
c 45	185	18.5	108317	12	US-10-017-161-2143	Sequence 2143, Ap

ALIGNMENTS

RESULT 1
US-10-275-910-1
; Sequence 1, Application US/10275910
; Publication No. US20030166142A1
; GENERAL INFORMATION:
; APPLICANT: Ramakrishnan, Shyam
; TITLE OF INVENTION: REGULATION OF HUMAN P2Y-LIKE G PROTEIN-COUPLED RECEPTOR
; FILE REFERENCE: 4974.03885
; CURRENT APPLICATION NUMBER: US/10/275,910
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/203,582
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/269,857
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: gene
; LOCATION: (520)..(1373)
US-10-275-910-1

Query Match 19.1%; Score 191.4; DB 12; Length 3143;
Best Local Similarity 88.8%; Pred. No. 2.1e-28;
Matches 207; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Qy 13 CATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAG 72
Db 2723 CATTATCTAGGCTGGGCAGAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCAAG 2782

Qy 73 GTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCT 132
Db 2783 GCGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAATCTGGCCAACATGGTGAAACCCC 2842

Qy 133 GTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCT 192

```
Db      2843 ATCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGGCGGGTGCCTATAATCCCAGCT 2902

Qy       193 ACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      2903 ACTCGGGAGGCTGAGAAAGGAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCA 2955
```

RESULT 2
US-09-764-891-6869
; Sequence 6869, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6869
; LENGTH: 30573
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6869

```
  Query Match             19.1%;  Score 191.4;  DB 11;  Length 30573;
  Best Local Similarity   90.7%;  Pred. No. 5.1e-28;
  Matches  204;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        21 AGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGG 80

Db      6984 AGGCTGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGTCGAGGTAGGTGG 7043

Qy        81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTAC 140

Db      7044 ATCATCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 7103

Qy       141 TAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGA 200

Db      7104 TAAAAATACAACAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTTGGGA 7163

Qy       201 GGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      7164 GGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCA 7208
```

RESULT 3
US-09-764-891-6868
; Sequence 6868, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6868
; LENGTH: 32195
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6868

```
  Query Match             19.1%;  Score 191.4;  DB 11;  Length 32195;
  Best Local Similarity   90.7%;  Pred. No. 5.3e-28;
  Matches  204;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy        21 AGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGGTGGGCGG 80

Db     12766 AGGCTGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGAGGTCGAGGTAGGTGG 12825

Qy        81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTGTCTCTAC 140

Db     12826 ATCATCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACCCTGTCTCTAC 12885

Qy       141 TAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTACTCGGGA 200

Db     12886 TAAAAATACAACAATTAGCTGGGCATGGTGGCACGTGCCTGTAATCCCAGCTACTTGGGA 12945

Qy       201 GGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db     12946 GGCTGAGGCAGGAGAATTGCTTGAACCTGGGAGGTGGAGGTTGCA 12990
```

RESULT 4
US-09-814-353-21993
; Sequence 21993, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21993
; LENGTH: 7677
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1, 2, 3, 7675, 7676, 7677
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21993

```
  Query Match             19.1%;  Score 190.8;  DB 12;  Length 7677;
  Best Local Similarity   60.7%;  Pred. No. 3.9e-28;
  Matches  368;  Conservative    0;  Mismatches  227;  Indels   11;  Gaps    3;

Qy         1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60

Db      6836 AATTTATTAAGAATTAATCTCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTT 6895

Qy        61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120

Db      6896 TGGGAAGCCAAGGTGGGCGGATCACTTCAGGTTAGGAGTTGGAGACCAGCCTGGCCAACA 6955

Qy       121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180

Db      6956 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCACTTGCCT 7015

Qy       181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240

Db      7016 GTAATCCCAGCTACTCGGGAGGCTGAGGCACGAGAATCGATTAAACCTGGGAGGCGGAGA 7075

Qy       241 TTGC----ATATCTGAGTGGTGAAATTGTGATTCTTTTTTTCTCTTTGTCTGTATTTTTG 296
```

```
Db       7076 TTGCAATGAGATCGAACCACTGCACTCCAGCCTGGGTGACAGAGAGAGACACTGCCTTGG 7135

Qy        297 AACTTTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTCAAAT 356

Db       7136 AAAAAAAAAGAATCTCACTCACTATCTAGAGAGGATTGTCAGAATATTCACGATTCAGGT 7195

Qy        357 GTTAATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTAAGGT 416

Db       7196 CTTGAAACTTTGATTATGCAAAAGAAGGTATATAATAAATATTTCATTATGATTCAGTTT 7255

Qy        417 TTCTTGTGTAGCAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGA 476

Db       7256 TTAAGGCTTTGCAGCT---TCTATAAGTGTTCTCAGATGCCACTAGATAATTTTAAAAGC 7312

Qy        477 ATACATACACGTATATTTTTTATAACATTA----AGAAACAGCAGCATCATTACTTTAAT 532

Db       7313 ATCATATTAGAAATACTTTAAGAAGACTTATATAAGAAATAGAAGATTGTTGAATTTTAC 7372

Qy        533 CCATCATTTCGTTAACCACCATATAGCTGTTGATCATTGTATTGTCATGTGCTTTTTAA 592

Db       7373 AGAGGATTTGGTTCATTAAGACCCAGATTCTGTAAGTTTTCATTCTGAAATTGTAGTTAA 7432

Qy        593 AAATCT 598

Db       7433 ACATAT 7438


RESULT 5
US-09-836-613-3
; Sequence 3, Application US/09836613
; Publication No. US20030039643A1
;  GENERAL INFORMATION:
;       APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;
;                  WEBER, BIRGIT; BLANCH, LIANNE; ANSON, DONALD STEWART
;       TITLE OF INVENTION: SYNTHETIC MAMMALIAN
;                           N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES
;                           ENCODING SAME
;       NUMBER OF SEQUENCES: 6
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: NIXON PEABODY LLP
;            STREET: 990 STEWART AVENUE
;            CITY: GARDEN CITY
;            STATE: NEW YORK
;            COUNTRY: UNITED STATES
;            ZIP: 11530
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/836,613
;            FILING DATE: 17-Apr-2001
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/US96/00747
;            FILING DATE: 22-NOV-1996
;       ATTORNEY/AGENT INFORMATION:
;            NAME: POKALSKY, ANN R.
;            REGISTRATION NUMBER: 34,697
;            REFERENCE/DOCKET NUMBER: 2249/104
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 516 742 4343
;            TELEFAX: 516 742 4366
;  INFORMATION FOR SEQ ID NO: 3:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 10380 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: DNA (genomic)
;       ORIGINAL SOURCE:
;            ORGANISM: Homo sapiens
;       POSITION IN GENOME:
;            CHROMOSOME/SEGMENT: Chromosome 17
;       FEATURE:
;            NAME/KEY:  exon 1
;            LOCATION:  990..1372
;       FEATURE:
;            NAME/KEY:  exon 2
;            LOCATION:  2115..2262
;       FEATURE:
;            NAME/KEY:  exon 3
;            LOCATION:  3056..3202
;       FEATURE:
;            NAME/KEY:  exon 4
;            LOCATION:  3387..3472
;       FEATURE:
;            NAME/KEY:  exon 5
;            LOCATION:  5667..5923
;       FEATURE:
;            NAME/KEY:  exon 6
;            LOCATION:  7745..8955
;       SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-836-613-3

  Query Match             19.0%;  Score 189.8;  DB 11;  Length 10380;
  Best Local Similarity   86.7%;  Pred. No. 7e-28;
  Matches  209;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64

Db       3540 TAAAAACTTAAGCTCTGGGCCGGGCGCAGTGGCTCACGCCTGTAATCCCAGCACTTTGGG 3599

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124

Db       3600 AGGCCGAGTTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGT 3659

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184

Db       3660 GAAACCCTGTCTCTACTAAAAATATAAAAATTAGCTGGGCATGGTGGCACGCGCCTGTAA 3719

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244

Db       3720 TCCCATCTACTTAGGAGGCTGAGACAGGAGAATTGCTTAAACCTGGGAGGCAGACGTTGC 3779

Qy        245 A 245

Db       3780 A 3780


RESULT 6
US-10-027-632-123771
; Sequence 123771, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123771
;   LENGTH: 1325
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-123771

  Query Match            18.9%;  Score 189.4;  DB 13;  Length 1325;
  Best Local Similarity  86.3%;  Pred. No. 3.8e-28;
  Matches  208;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          4 ATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGG 63
Db        206 ATTAAAAAAAAWTTTTTTGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGG 265

Qy         64 GAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGG 123
Db        266 GAGGCCAAGGTGGGTGGATTACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACCTGG 325

Qy        124 CGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTA 183
Db        326 TGAAACCCTGCATCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGGTGCCTGTA 385

Qy        184 ATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTG 243
Db        386 ATCCCAGCTACTCAGGAGGCTGAGGCAGGAAAATCACTTGAACCTGGGAGGCGGAGGTTG 445

Qy        244 C 244
              |
Db        446 C 446

RESULT 7
US-10-027-632-123772
; Sequence 123772, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123772
;   LENGTH: 1325
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-123772

  Query Match            18.9%;  Score 189.4;  DB 13;  Length 1325;
  Best Local Similarity  86.3%;  Pred. No. 3.8e-28;
  Matches  208;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          4 ATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGG 63
Db        206 ATTAAAAAAAAWTTTTTTGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGG 265

Qy         64 GAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGG 123
Db        266 GAGGCCAAGGTGGGTGGATTACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACCTGG 325

Qy        124 CGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTA 183
Db        326 TGAAACCCTGCATCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGGTGCCTGTA 385

Qy        184 ATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTG 243
Db        386 ATCCCAGCTACTCAGGAGGCTGAGGCAGGAAAATCACTTGAACCTGGGAGGCGGAGGTTG 445

Qy        244 C 244
              |
Db        446 C 446

RESULT 8
US-10-027-632-123773
; Sequence 123773, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 123773
;   LENGTH: 1325
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-123773

  Query Match            18.9%;  Score 189.4;  DB 13;  Length 1325;
  Best Local Similarity  86.3%;  Pred. No. 3.8e-28;
  Matches  208;  Conservative    1;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          4 ATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGG 63
Db        206 ATTAAAAAAAAWTTTTTTGGCTGGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGG 265

Qy         64 GAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGG 123
Db        266 GAGGCCAAGGTGGGTGGATTACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACCTGG 325

Qy        124 CGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTA 183
Db        326 TGAAACCCTGCATCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGGTGCCTGTA 385

Qy        184 ATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTG 243
Db        386 ATCCCAGCTACTCAGGAGGCTGAGGCAGGAAAATCACTTGAACCTGGGAGGCGGAGGTTG 445

Qy        244 C 244
              |
Db        446 C 446
```

```
RESULT 9
US-10-027-632-102144/c
; Sequence 102144, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,216
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 102144
;   LENGTH: 2490
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-102144

  Query Match             18.9%;  Score 189.4;  DB 13;  Length 2490;
  Best Local Similarity   87.0%;  Pred. No. 4.8e-28;
  Matches  208;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db        476 AAAAAAAAAAAAAAAAGGTCGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 417

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db        416 GGCCGAGGTGAGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 357

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db        356 AAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCGCATGCCTGTAAT 297

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
Db        296 CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCTTGAACCTGGGAGGCGGAGGTTGC 238

RESULT 10
US-10-027-632-111607
; Sequence 111607, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 111607
;   LENGTH: 2490
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-111607

  Query Match             18.9%;  Score 189.4;  DB 13;  Length 2490;
  Best Local Similarity   87.0%;  Pred. No. 4.8e-28;
  Matches  208;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db       2015 AAAAAAAAAAAAAAAAGGTCGGGCATGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGA 2074

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db       2075 GGCCGAGGTGAGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 2134

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db       2135 AAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCGCATGCCTGTAAT 2194

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244
Db       2195 CCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGGCTTGAACCTGGGAGGCGGAGGTTGC 2253

RESULT 11
US-10-027-632-306904
; Sequence 306904, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 306904
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-306904

  Query Match             18.9%;  Score 188.8;  DB 13;  Length 614;
  Best Local Similarity   88.4%;  Pred. No. 3.7e-28;
  Matches  205;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy         14 ATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGG 73
Db        244 ATATTTAAGGCTGGGCCTGGTGGCTCACTCCTGTAATCCCAGCACTTTGGGAGGCCAAGG 303
```

```
Qy          74 TGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTG 133
Db         304 CGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCTG 363

Qy         134 TCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTA 193
Db         364 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCCTGTAGTCCCAGCTA 423

Qy         194 CTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db         424 CTCAGGAGGCTGAGACAGGAGAATTGTTTGAACCAGGGAGGCGGAGGTTGCA 475
```

RESULT 12
US-10-027-632-306905
; Sequence 306905, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 306905
;   LENGTH: 614
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-306905

```
  Query Match             18.9%;  Score 188.8;  DB 13;  Length 614;
  Best Local Similarity   88.4%;  Pred. No. 3.7e-28;
  Matches  205;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          14 ATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGG 73
Db         244 ATATTTAAGGCTGGGCCTGGTGGCTCACTCCTGTAATCCCAGCACTTTGGGAGGCCAAGG 303

Qy          74 TGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTG 133
Db         304 CGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCTG 363

Qy         134 TCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTA 193
Db         364 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCCTGTAGTCCCAGCTA 423

Qy         194 CTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db         424 CTCAGGAGGCTGAGACAGGAGAATTGTTTGAACCAGGGAGGCGGAGGTTGCA 475
```

RESULT 13
US-10-027-632-77917
; Sequence 77917, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 77917
;   LENGTH: 944
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(944)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-77917

```
  Query Match             18.9%;  Score 188.8;  DB 13;  Length 944;
  Best Local Similarity   88.4%;  Pred. No. 4.3e-28;
  Matches  205;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          14 ATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGAGGCCAAGG 73
Db         244 ATATTTAAGGCTGGGCCTGGTGGCTCACTCCTGTAATCCCAGCACTTTGGGAGGCCAAGG 303

Qy          74 TGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCTG 133
Db         304 CGGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGACCAACATGGTGAAACCCTG 363

Qy         134 TCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAATCCCAGCTA 193
Db         364 TCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTGGCGTGCGCCTGTAGTCCCAGCTA 423

Qy         194 CTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db         424 CTCAGGAGGCTGAGACAGGAGAATTGTTTGAACCAGGGAGGCGGAGGTTGCA 475
```

RESULT 14
US-09-764-869-1599
; Sequence 1599, Application US/09764869
; Patent No. US20020061521A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC007
; CURRENT APPLICATION NUMBER: US/09/764,869
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2442
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1599
;   LENGTH: 17792
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-869-1599

```
  Query Match             18.9%;  Score 188.8;  DB 9;  Length 17792;
  Best Local Similarity   86.7%;  Pred. No. 1.4e-27;
  Matches  208;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;
```

```
Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65

Db      13401 AAAAAATCACTCTTTCGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 13460

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125

Db      13461 GGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 13520

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185

Db      13521 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCACACACCTGTAAT 13580

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      13581 CCCAGCTACTTGGGAGGCTGAGGCAAGAGAATAGCTTGAACCTGGGAGATGGAGGTTGCA 13640


RESULT 15
US-10-091-504-1599
; Sequence 1599, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC007C1
;  CURRENT APPLICATION NUMBER: US/10/091,504
;  CURRENT FILING DATE:  2002-03-07
;  NUMBER OF SEQ ID NOS: 2442
;  Prior Application removed - See File Wrapper or Palm
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1599
;   LENGTH: 17792
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-091-504-1599

  Query Match           18.9%;  Score 188.8;  DB 14;  Length 17792;
  Best Local Similarity   66.7%;  Pred. No. 1.4e-27;
  Matches  209;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65

Db      13401 AAAAAATCACTCTTTCGGCTGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGA 13460

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125

Db      13461 GGCCAAGGTGGGCGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 13520

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185

Db      13521 AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGGTGGCACACACCTGTAAT 13580

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245

Db      13581 CCCAGCTACTTGGGAGGCTGAGGCAAGAGAATAGCTTGAACCTGGGAGATGGAGGTTGCA 13640


Search completed: October 24, 2003, 18:34:54
Job time : 250.585 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 2091.3 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....acctgattctggatcttcttt 1001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	530.4	53.0	556	14	CB241186	CB241186 UI-CF-FN0
2	511.2	51.1	532	9	AW297897	AW297897 UI-H-BW0-
3	403	40.3	743	13	BU853405	BU853405 AGENCOURT
4	398.2	39.8	877	13	BU193776	BU193776 AGENCOURT
5	388.6	38.8	754	10	BG204678	BG204678 RST24695
6	388	38.8	749	10	BG217901	BG217901 RST37625
c 7	350	35.0	894	10	BG193969	BG193969 RST13111
8	286.2	28.6	729	10	BF216343	BF216343 602884376
9	277.8	27.8	360	10	BG185583	BG185583 RST4534 A
c 10	244.6	24.4	460	10	BE172740	BE172740 MR0-HT0
11	218.2	21.8	239	10	BG208384	BG208384 RST27882
12	209.4	20.9	401	14	W42916	W42916 zc25f07.r1
c 13	193.4	19.3	516	28	AQ112451	AQ112451 CIT-HSP-2
c 14	188.8	18.9	515	14	CA946732	CA946732 is09h12.x
c 15	187.4	18.7	446	9	AI339725	AI339725 qq42e08.x
16	187.4	18.7	469	10	BG574581	BG574581 602596564
17	187.4	18.7	660	9	AL708207	AL708207 DKFZp686L
c 18	187.2	18.7	660	28	AQ057239	AQ057239 CIT-HSP-2
19	187.2	18.7	711	28	AQ415030	AQ415030 RPCI-11-2
20	187	18.7	499	2	HSM096977	Bx503260 Homo sapi
c 21	186.8	18.7	363	9	AA493808	AA493808 nh02f06.s
c 22	186.6	18.6	500	9	AW338860	AW338860 ha67f11.x
c 23	186.6	18.6	826	13	BQ438665	BQ438665 AGENCOURT
c 24	186.4	18.6	400	2	HSM095126	Bx492298 Homo sapi
25	186.4	18.6	410	28	AQ673107	AQ673107 HS_5333_B
c 26	186.4	18.6	794	12	BM679479	BM679479 UI-E-EO0
27	186.2	18.6	368	9	AA603264	AA603264 np16c01.s
28	186.2	18.6	597	9	AI821805	AI821805 nr27a08.x
c 29	186.2	18.6	407	9	AI821056	AI821056 nr27a08.y
c 30	186.2	18.6	425	9	AI792092	AI792092 np16c01.y
31	186.2	18.6	439	9	AI791718	AI791718 oq53c06.y
c 32	186.2	18.6	442	28	AQ018157	AQ018157 CIT-HSP-2
33	186.2	18.6	600	13	BX093702	BX093702 BX093702
34	186.2	18.6	707	28	AQ540344	AQ540344 RPCI-11-3
35	185.8	18.6	544	9	AL699310	AL699310 DKFZp686C
c 36	185.6	18.5	443	10	BF991881	BF991881 PM4-GN001
c 37	185.6	18.5	583	28	AQ391420	AQ391420 CITBI-E1-
38	185.6	18.5	662	29	AG037584	AG037584 Pan trogl
39	185.4	18.5	361	9	AW265688	AW265688 xq74g11.x
40	185.4	18.5	700	28	AQ317190	AQ317190 RPCI1-10
41	185.4	18.5	735	29	AG094268	AG094268 Pan trogl
c 42	185.2	18.5	435	10	BF991882	BF991882 PM4-GN001
43	185.2	18.5	607	29	BZ603705	BZ603705 WHADP21TR
44	185	18.5	385	9	AA278496	AA278496 zs81d04.r
c 45	184.8	18.5	342	9	AA846923	AA846923 oe08d01.s

ALIGNMENTS

RESULT 1
CB241186
LOCUS CB241186 556 bp mRNA linear EST 12-FEB-2003
DEFINITION UI-CF-FN0-afx-b-20-0-UI.s1 UI-CF-FN0 Homo sapiens cDNA clone
UI-CF-FN0-afx-b-20-0-UI 3', mRNA sequence.
ACCESSION CB241186
VERSION CB241186.1 GI:28362830
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 556)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa

```
         cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
         cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
         DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
         Clone Distribution: Researchers may obtain clones from Research
         Genetics (www.resgen.com) or from Open Biosystems
         (www.openbiosystems.com).
         The following repetitive elements were found in this cDNA
         sequence: 1-22, >AT_rich#Low_complexity (matched compliment)
         Seq primer: M13 FORWARD
         POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..556
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-CF-FN0-afx-b-20-0-UI"
                     /tissue_type="Human Lung Epithelial cells"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-CF-FN0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-CF-FN0 is a subtracted cDNA library derived from two
                     normalized Human lung epithelial cell libraries (EN1 and
                     DU1) The library was subtracted according to according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. For additional information, contact:
                     bento-soares@uiowa.edu
                     TAG_LIB=UI-CF-FN0
                     TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
                     TAG_SEQ=GGCTGTAGGC"
BASE COUNT      196 a     93 c     80 g    187 t
ORIGIN

  Query Match             53.0%;  Score 530.4;  DB 14;  Length 556;
  Best Local Similarity   99.8%;  Pred. No. 5.5e-07;
  Matches  531;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy        391 TATATTAGTATAGCATTTATTAAGGTTTCTTGTGTAGCAGATCAACATAGAAAATATATT 450
Db         16 TATATTAGTATAGCATTTATTAAGGTTTCTTGTGTAGCAGATCAACATAGAAAATATATT 75

Qy        451 TAAATGGCTGACATAATTTTCTAAGAATACATACACGTATATTTTTTATAACATTAAGAA 510
Db         76 TAAATGGCTGACATAATTTTCTAAGAATACATACACGTATATTTTTTATAACATTAAGAA 135

Qy        511 ACAGCAGCATCATTACTTTAATCCATCATTTCGTTAACCACCATATACCTGTTGATCATT 570
Db        136 ACAGCAGCATCATTACTTTAATCCATCATTTCGTTAACCACCATATACCTGTTGATCATT 195

Qy        571 TGTATTGTCATGTGCTTTTTAAAAATCTAGATGAGAAATATTCGATTATCTGACTTCACT 630
Db        196 TGTATTGTCATGTGCTTTTTAAAAATCTAGATGAGAAATATTCGATTATCTGACTTCACT 255

Qy        631 GAATCCTTGAAAAAAATAAAACGCAGCGTCAGCCCTCAAACTTTAGAAGCGTACATACGT 690
Db        256 GAATCCTTGAAAAAAATAAAACGCAGCGTCAGCCCTCAAACTTTAGAAGCGTACATACGT 315

Qy        691 TGGAACAAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACCTGCAGAAC 750
Db        316 TGGAACAAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACCTGCAGAAC 375

Qy        751 ATTTTACTTAAAAGAGGAAACACAAGATCTTCAATGAACGTCATCGGCTACAGAAACAGC 810
Db        376 ATTTTACTTAAAAGAGGAAACACAAGATATTCAATGAACGTCATCGGCTACAGAAACAGC 435

Qy        811 CTAAGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAACC 870
Db        436 CTAAGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAACC 495

Qy        871 AGAAAACAGACTTAAACAAAATATACAATGCAAATGTAATTTTTTGTTGTTT 922
Db        496 AGAAAACAGACTTAAACAAAATATACAATGCAAATGTAATTTTTTGTTGTTT 547


RESULT 2
AW297897
LOCUS       AW297897                 532 bp    mRNA    linear   EST 16-JAN-2000
DEFINITION  UI-H-BW0-aju-f-11-0-UI.s1 NCI_CGAP_Sub6 Homo sapiens cDNA clone
            IMAGE:2733092 3', mRNA sequence.
ACCESSION   AW297897
VERSION     AW297897.1  GI:6704522
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 532)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            NCI-CGAP clone distribution information can be found through the
            I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html The following repetitive
            elements were found in this cDNA sequence: 1-23,
            >AT_rich#Low_complexity
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..532
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2733092"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Sub6"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; NCI_CGAP_Sub6
                     is a subtracted library derived from BW, which consists of
                     a mixture of four normalized libraries: NCI_CGAP_Brn50,
                     NCI_CGAP_Lu13, NCI_CGAP_Ov18, GBC1 . The NCI_CGAP_Sub6
                     library had 7 million recombinants. A single-stranded DNA
                     preparation of BW was used as a tracer in a subtractive
                     hybridization with a driver comprising: the IMAGE pool
                     (NCI_CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
                     3798-3803 (IMAGE CloneIDs 1322376-1323911,
                     1456008-1456775,1500552-1502855); NCI_CGAP_Kid5 pool 1
                     LLAM 3338-3342,3722-3725, 3776-3778 (IMAGE CloneIDs
                     1323912-1325831, 1471368-1472903, 1492104-1493255);
                     NCI_CGAP_Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
                     CloneIDs 1414920-1417991, 1520904-1522439); NCI_CGAP_GC4
                     pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
                     CloneIDs 1257096-1258631,1469064-1470983, 1475592-1476743
                     ); NCI_CGAP_Pr22 pool 1 LLAM 2457-2459, 2758-2759,
                     3062-3068 (IMAGE CloneIDs 985608-986759, 1101192-1101959,
                     1217928-1220615); NCI_CGAP_Col0 pool 1 LLAM 2644-2653,
                     2871-2872 (IMAGE CloneIDs 1057416-1061255,1144584-1145351
                     ). (50% of the driver population), plus a pool of 3,840
                     arrayed clones from NCI_CGAP_Sub1 (IMAGE CloneIDs
                     2708616-2710535) and NCI_CGAP_Sub2 (IMAGE CloneIDs
                     2710536-2712455)    (20% of the driver population), plus a
                     pool of 11,136 clones from NCI_CGAP_Sub3 (IMAGE CloneIDs
                     2712456-2723591) (30% of the driver population).
                     Subtraction was performed as previously described [Bonaldo
                     , Lennon & Soares (1996): Normalization and Subtraction:
                     Two Approaches To Facilitate Gene Discovery. Genome
                     Research 6, 791-806.
                     TAG_LIB=NCI_CGAP_Brn50
                     TAG_TISSUE=brain
                     TAG_SEQ=TTTCG"
```

```
BASE COUNT      194 a     89 c     77 g    172 t
ORIGIN

  Query Match           51.1%;  Score 511.2;  DB 9;  Length 532;
  Best Local Similarity   99.4%;  Pred. No. 1.9e-06;
  Matches  513;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        391 TATATTAGTATAGCATTTATTAAGGTTTCTTGTGTAGCAGATCAACATAGAAAATATATT 450
              |||||| | |||| | |||| ||| ||| ||| || ||| |
Db         17 TAAATAAGTATAGCATTTATTAAGGTTTCTTGTGTAGCAGATCAACATAGAAAATATATT  76

Qy        451 TAAATGGCTGACATAATTTTCTAAGAATACATACACGTATATTTTTTATAACATTAAGAA 510
              |||||| | |||||||||||| |||| ||| ||||||||||| || ||| |
Db         77 TAAATGGCTGACATAATTTTCTAAGAATACATACACGTATATTTTTTATAACATTAAGAA 136

Qy        511 ACAGCAGCATCATTACTTTAATCCATCATTTCGTTAACCACCATATACCTGTTGATCATT 570
              |||||| |||| ||| ||| ||| |||||||| ||| ||| || |
Db        137 ACAGCAGCATCATTACTTTAATCCATCATTTCGTTAACCACCATATACCTGTTGATCATT 196

Qy        571 TGTATTGTCATGTGCTTTTTAAAAATCTAGATGAGAAATATTCGATTATCTGACTTCACT 630
              |||||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||
Db        197 TGTATTGTCATGTGCTTTTTAAAAATCTAGATGAGAAATATTCGATTATCTGACTTCACT 256

Qy        631 GAATCCTTGAAAAAAAATAAAACGCAGCGTCAGCCCTCAAACTTTAGAAGCGTACATACGT 690
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 GAATCCTTGAAAAAAATAAAACGCAGCGTCAGCCCTCAAACTTTAGAAGCGTACATACGT 316

Qy        691 TGGAACAAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACCTGCAGAAC 750
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 TGGAACAAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACCTGCAGAAC 376

Qy        751 ATTTTACTTAAAAGAGGAAACACAAGATCTTCAATGAACGTCATCGGCTACAGAAACAGC 810
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 ATTTTACTTAAAAGAGGAAACACAAGATCTTCAATGAACGTCATCGGCTACAGAAACAGC 436

Qy        811 CTAAGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAACC 870
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CTAAGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAGCC 496

Qy        871 AGAAAACAGACTTAAACAAAATATACAATGCAAATG 906
              ||||||||||||||||||||||||||||||||||
Db        497 AGAAAACAGACTTAAACAAAATATACAATGCAAATG 532


RESULT 3
BU853405
LOCUS       BU853405                 743 bp    mRNA    linear   EST 16-OCT-2002
DEFINITION  AGENCOURT_10418018 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6620245
            5', mRNA sequence.
ACCESSION   BU853405
VERSION     BU853405.1  GI:24038371
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 743)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2870  row: a  column: 13
            High quality sequence stop: 648.
FEATURES             Location/Qualifiers
     source          1..743
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6620245"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_82"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
BASE COUNT      248 a    123 c    121 g    250 t      1 others
ORIGIN

  Query Match           40.3%;  Score 403;  DB 13;  Length 743;
  Best Local Similarity   100.0%;  Pred. No. 0.0012;
  Matches  403;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
              |||||||||||| ||| ||| |||| |||||  ||||| | ||||| ||||
Db         11 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG  70

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
              |||| ||| ||| |||| |||| |||| |||| ||||| ||||| |||||| |||||
Db         71 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 130

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
              |||| ||| |||| ||| |||| |||| |||| | |||||  |||||
Db        131 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 190

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 250

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
              |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 310

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        311 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 370

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
              |||||||||||||||||||||||||||||||||||||||||||
Db        371 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 413


RESULT 4
BU193776
LOCUS       BU193776                 877 bp    mRNA    linear   EST 04-SEP-2002
DEFINITION  AGENCOURT_7905132 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6161142
            5', mRNA sequence.
ACCESSION   BU193776
VERSION     BU193776.1  GI:22707760
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 877)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
```

```
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13512  row: p  column: 07
            High quality sequence stop: 579.
FEATURES             Location/Qualifiers
     source          1..877
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6161142"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      300 a    166 c    177 g    234 t
ORIGIN

  Query Match             39.8%;  Score 396.2;  DB 13;  Length 877;
  Best Local Similarity   99.3%;  Pred. No. 0.0015;
  Matches  400;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
Db        293 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 352

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
Db        353 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 412

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
Db        413 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 472

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 838
Db        473 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTT 532

Qy        839 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 898
Db        533 ACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAA 592

Qy        899 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
Db        593 TGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGG 652

Qy        959 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db        653 ACACTAAGTTAGAGCACAACAAAACCTGATTCTGGGCTTCCTT 695


RESULT 5
BG204678
LOCUS       BG204678              784 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST24093 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG204678
VERSION     BG204678.1  GI:13726365
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 784)
  AUTHORS   Harrington,J.C., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
   PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 500.
FEATURES             Location/Qualifiers
     source          1..784
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      255 a    141 c    129 g    259 t
ORIGIN

  Query Match             38.8%;  Score 388.6;  DB 10;  Length 784;
  Best Local Similarity   99.0%;  Pred. No. 0.0029;
  Matches  391;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy        607 AATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTCAGCCCT 666
Db          7 AATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTCAGCCCT 66

Qy        667 CAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTTTAAGGA 726
Db         67 CATACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTTTAAGGA 126

Qy        727 AATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCTTCAATG 786
Db        127 AATACCTTTGTAAACCTGCAGAACACTTTACTTAAAAGAGGAAACACAAGATCTTCAATG 186

Qy        787 AACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTACATATTT 846
Db        187 AACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTACATATTT 246

Qy        847 GTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATGCAAATG 906
Db        247 GTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTATACAAAATATACAATGCAAATG 306

Qy        907 TAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGACACTAAG 966
Db        307 TAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGCTATCCCAATGGACACTAAG 366

Qy        967 TTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db        367 TTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 401


RESULT 6
BG217901
LOCUS       BG217901              749 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST37625 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG217901
VERSION     BG217901.1  GI:13743922
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 749)
  AUTHORS   Harrington,J.C., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
```

```
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
   PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 551.
FEATURES             Location/Qualifiers
     source          1..749
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      251 a    127 c    121 g    250 t
ORIGIN

  Query Match             38.8%;  Score 388;  DB 10;  Length 749;
  Best Local Similarity   99.8%;  Pred. No. 0.0031;
  Matches  399;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy        603 GAGAAATATTCGATTATC TGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTCA 661
Db          1 GAGAAATATTCGATTATCGTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCGTCA 60

Qy        662 GCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTTT 721
Db         61 GCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTGTTT 120

Qy        722 AAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCTT 781
Db        121 AAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGATCTT 180

Qy        782 CAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTACA 841
Db        181 CAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCTTACA 240

Qy        842 TATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATGC 901
Db        241 TATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACAATGC 300

Qy        902 AAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGACA 961
Db        301 AAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATGGACA 360

Qy        962 CTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db        361 CTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 400

RESULT 7
BG193969/c
LOCUS       BG193969                 894 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST13111 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG193969
VERSION     BG193969.1  GI:13715656
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 894)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
   PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 461.
FEATURES             Location/Qualifiers
     source          1..894
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      284 a    154 c    169 g    283 t      4 others
ORIGIN

  Query Match             35.0%;  Score 350;  DB 10;  Length 894;
  Best Local Similarity   96.3%;  Pred. No. 0.029;
  Matches  389;  Conservative    0;  Mismatches   12;  Indels    3;  Gaps    3;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
Db        770 AGATGAGAAAAATTCGATAATCTGACTTCANTGAATCCTTGAAAAAATTAAAACGCAGCG 711

Qy        659 TCAGCCCTCAAACTTT-AGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACT 717
Db        710 TCAGCCTTCTAACTTTAAGAAGCGGACATACGTTGGACCAAGGACTTTGGAGATACCACT 651

Qy        718 GTTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGA 777
Db        650 GTTTAAGNAAAT-CCTTTGTTAACCTGCAGAACATTTT-CTTAAAAGAGGAAACACAAGA 593

Qy        778 TCTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGTCT 837
Db        592 TCTTCAATGAACGTCATCGGCTACAGAAACAGCTTAAGTTTACAGGACTTTTTAGAGTCT 533

Qy        838 TACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACA 897
Db        532 TACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATACA 473

Qy        898 ATGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATG 957
Db        472 ATGCCAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAATG 413

Qy        958 GACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db        412 GACACTAAGTTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 369

RESULT 8
BF216343
LOCUS       BF216343                 728 bp    mRNA    linear   EST 06-NOV-2000
```

```
DEFINITION  601884376F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4102722 5',
            mRNA sequence.
ACCESSION   BF216343
VERSION     BF216343.1  GI:11109929
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 728)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM976  row: h  column: 19
            High quality sequence stop: 574.
FEATURES             Location/Qualifiers
     source          1..728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4102722"
                     /tissue_type="glioblastoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_57"
                     /note="Organ: brain; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows. 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.55 kb (range 0.9-4.0 kb).  12/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      248 a    132 c    150 g    198 t
ORIGIN

  Query Match             28.6%;  Score 286.2;  DB 10;  Length 728;
  Best Local Similarity   92.6%;  Pred. No. 1.8;
  Matches  377;  Conservative    0;  Mismatches   23;  Indels    7;  Gaps    7;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 656
Db        142 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAG-G 200

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTT-TGGAGATACCACT 717
Db        201 TCAGCCCTCAAA-TTTAGAAGCGTACATACGTTGGAACAAGGACTTGTGGAGATACCACT 259

Qy        718 -GTTTAAGGAAATACCTT-TGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAA 775
Db        260 GGTTTAAGGAAATACCTTGTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAA 319

Qy        776 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGT 835
Db        320 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGT 379

Qy        836 CTTACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAACAAAATATA 895
Db        380 CTTACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGA-TTAAACAAAATATA 438

Qy        896 CAATGCAAATGTAATTTTTTGTTGTTTAAGGCCTTGCCTTGATGGTCACAGTTATCCCAA 955
Db        439 CAATGCAAATGTAATTTTGGTGGTCTAAGGCCTTGCTTGGGATGGTCACAGTTATCCCAA 498

Qy        956 TGGACACTAAG-TTAGAGCACAACAAAACCTGATTCTGGTCTTCTTT 1001
Db        499 TGGACACTAAGTTTAGAGCACACAAAAACTGATTTCTGGTCTTCTTT 545


RESULT 9
BG185583
LOCUS       BG185583                 360 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST4534 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG185583
VERSION     BG185583.1  GI:13707270
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 360)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
   PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 310.
FEATURES             Location/Qualifiers
     source          1..360
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT      129 a     61 c     63 g    107 t
ORIGIN

  Query Match             27.8%;  Score 277.8;  DB 10;  Length 360;
  Best Local Similarity   97.1%;  Pred. No. 5;
  Matches  304;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    2;

Qy        689 GTTGGAACAAGGACTTTGGAGATACCACTGTTTAAGGAAATACCTTTGTAAACCTGCAGA 748
Db         15 GTTGGAACAAGGACTTTGGAGAT-CCACTGTTGAAGGAAATACCATTGTAAACCTGCAGA 73

Qy        749 ACATTTTACTTAAAAGAGGAAACACAAGATCTTCAATGAACGTCATCGGCTACAGAAACA 808
Db         74 ACATTTTACTTAAAAGAGGAAACACAAGATCTTTAATGAGCGTGATCGGTTACAGAAACA 133

Qy        809 GCCTAAGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAA 868
Db        134 GCCT-AGTTTACAGGACTTTTTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAA 192

Qy        869 CCAGAAAACAGACTTAAACAAAATATACAATGCAAATGTAATTTTTTGTTGTTTAAGGCC 928
Db        193 CCAGAAAACAGACTTAAACAAAATATACAGTGCAAATGTAATTTTTTGTTGTTTAAGGCC 252

Qy        929 TTGCCTTGATGGTCACAGTTATCCCAATGGACACTAAGTTAGAGCACAACAAAACCTGAT 988
```

```
Db        253 TTGCCTTGATGGTCACAGTTATCCCAATGGACACTAAGTTAGAGCACAACAAAACCTGAT 312

Qy        989 TCTGGTCTTCTTT 1001

Db        313 TCTGGTCTTCTTT 325


RESULT 10
BE172740/c
LOCUS       BE172740                 460 bp    mRNA    linear   EST 21-JUN-2000
DEFINITION  MR0-HT0559-120400-010-a09 HT0559 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BE172740
VERSION     BE172740.1  GI:8635466
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 460)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR0-HT0559-120
            400-010-a09&t3=2000-04-12&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 6
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1..460
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0559"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      117 a    112 c     84 g    147 t
ORIGIN

  Query Match             24.4%;  Score 244.6;  DB 10;  Length 460;
  Best Local Similarity   74.6%;  Pred. No. 33;
  Matches  326;  Conservative    0;  Mismatches   99;  Indels   12;  Gaps    1;

Qy          5 TAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGG 64

Db        460 TAATAATAATAAAAAAAGGCCAGGTGCGGTGGTTCACACCTGTAATCCCAGCACTTTGGG 401

Qy         65 AGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124

Db        400 AGGCCACGAGACAGGTGGATCATGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACGTGGT 341

Qy        125 GAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 184

Db        340 GAAACCCTGCCTCTACTAAAAACACAAAAATTAGCTAGGCGTGGTGGCGGGAGGCTGTAA 281

Qy        185 TCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGC 244

Db        280 TCCCAGCTACTAGGGAGGTTGAGGCAGGAGAATCACTTGAATCTGGAAGGCAGAAGTTGC 221

Qy        245 ATATCTGAGTGGTGAAATTGTGATTCT-----------TTTTTTCTCTTTGTCTGTATT 292

Db        220 AGTGAGCCGAGGTCACACCATTGCACTCCAGCCTGGGCGACAAGAGTGAAACTCTGTCTC 161

Qy        293 TTTGAACTTTTCTATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTC 352

Db        160 AAAAAATATATATATAAATGATTGTGTTTTGTTTTTATATTGGAAAAATATTATGCTTTC 101

Qy        353 AAATGTTAATACCTATGAAACTAAACACAAGTAATAAATATATTAGTATAGCATTTATTA 412

Db        100 AAATGTTAATACCTATGAAACTAGACACAAGTAATAAATATATTAGTATAGCATTTATTA 41

Qy        413 AGGTTTCTTGTGTAGCA 429

Db         40 AGGTTTCTTGTGTAGCA 24


RESULT 11
BG208384
LOCUS       BG208384                 239 bp    mRNA    linear   EST 21-APR-2001
DEFINITION  RST27882 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION   BG208384
VERSION     BG208384.1  GI:13730071
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 239)
  AUTHORS   Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
            Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.
            , Lerner,L., Costanzo,D., McElligott,K., Boozer,S., Mays,R., Smith
            ,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
            ,J., Danzig,J. and Ducar,M.
  TITLE     Creation of genome-wide protein expression libraries using random
            activation of gene expression
  JOURNAL   Nat. Biotechnol. 19 (5), 440-445 (2001)
  MEDLINE   21227151
   PUBMED   11329013
COMMENT     Contact: Scott J. Cain
            Athersys, Inc.
            3201 Carnegie Ave, Cleveland, OH 44115, USA
            Tel: 216 431 9900
            Fax: 216 361 9596
            Email: scain@athersys.com
            High quality sequence stop: 189.
FEATURES             Location/Qualifiers
     source          1..239
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_line="HT1080"
                     /clone_lib="Athersys RAGE Library"
                     /note="See 'Creation of Genome-wide Protein Expression
                     Libraries using Random Activation of Gene Expression',
                     Nature Biotechnology, in press. Note that even though the
                     cell type indicated is HT1080, since a random activation
                     method was used, these sequence tags are not necessarily
                     expressed in HT1080 under normal circumstances."
BASE COUNT       88 a     46 c     42 g     63 t
ORIGIN

  Query Match             21.8%;  Score 218.2;  DB 10;  Length 239;
```

```
  Best Local Similarity   96.5%;  Pred. No. 2.8e+02;
  Matches  223;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        656 GCGTCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCA 715
Db          8 GTGTAAGCCCTCATTTTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCA 67

Qy        716 CTGTTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAA 775
Db         68 CTGTTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAA 127

Qy        776 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGT 835
Db        128 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTTAGAGT 187

Qy        836 CTTACATATTTGTGCACCAAACTTGAAGATGAACCAGAAAACAGACTTAAA 886
Db        188 CTAACATATTTGTGCACCAAACTTGAAGATGAACCATATAACAGACTTAAA 236


RESULT 12
W42916
LOCUS       W42916                   401 bp    mRNA    linear   EST 10-OCT-1996
DEFINITION  zc25f07.r1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA
            clone IMAGE:323365 5' similar to SW:YEN7_YEAST P39955 HYPOTHETICAL
            100.3 KD PROTEIN IN MEI4-CAJ1 INTERGENIC REGION. [1] ;, mRNA
            sequence.
ACCESSION   W42916
VERSION     W42916.1  GI:1327407
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2550   Std Error: 0.00
            Seq primer: mob.REGA+ET
            High quality sequence stop: 291.
FEATURES             Location/Qualifiers
     source          1..401
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:1254877"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:323365"
                     /tissue_type="senescent fibroblast"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares_senescent_fibroblasts_NbHSF"
                     /note="Vector: pT7T3D (Pharmacia) with a modified
                     polylinker V_TYPE: phagemid; Site_1: Not I; Site_2: Eco RI
                     ; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCATTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). Library went through one round of
                     normalization to a Cot = 5. Library constructed by Bento
                     Soares and M.Fatima Bonaldo."
BASE COUNT      142 a     76 c     78 g     95 t     10 others
ORIGIN

  Query Match             20.9%;  Score 209.4;  DB 14;  Length 401;
  Best Local Similarity   96.3%;  Pred. No. 3.3e+02;
  Matches  210;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy        599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAAATAAAACGCAGCG 658
Db        183 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGANNANAATAAAACGCAGCG 242

Qy        659 TCAGCCCTCAAACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 718
Db        243 TCAGCCTTCANACTTTAGAAGCGTACATACGTTGGAACAAGGACTTTGGAGATACCACTG 302

Qy        719 TTTAAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAAGAGGAAACACAAGAT 778
Db        303 TTTAAGGAAATACCTTTGTAAACCTGCNGAACATTTTACTTANAAGAGGAAACACAAGAT 362

Qy        779 CTTCAATGAACGTCATCGGCTACAGAAACAGCCTAAGT 816
Db        363 CTTCANTGAACGTCATCGGCTACAGAAACAGCCTAAGT 400


RESULT 13
AQ112451/c
LOCUS       AQ112451                 516 bp    DNA     linear   GSS 29-AUG-1998
DEFINITION  CIT-HSP-2372C9.TF CIT-HSP Homo sapiens genomic clone 2372C9,
            genomic survey sequence.
ACCESSION   AQ112451
VERSION     AQ112451.1  GI:3484611
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 516)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CIT-HSP-2372C9.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..516
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2372C9"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT       87 a    145 c    118 g    166 t
ORIGIN

  Query Match             19.0%;  Score 190.4;  DB 28;  Length 516;
  Best Local Similarity   87.1%;  Pred. No. 8.8e-02;
  Matches  209;  Conservative    0;  Mismatches   31;  Indels    0;  Gaps    0;
```

```
Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db        506 AAAAATGCAAAAATTAGGCTGGGCGCAGTGGCCCAGGCCTGTAATCCCAGCACTTTGGGA 447

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db        446 GGCCAAGGTGGGTGGATCACCTAAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 387

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db        386 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGTGTGCGCCTGTAAT 327

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db        326 CCCAGCTACTTGGGAGGCTGAGACAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCA 267


RESULT 14
CA946732/c
LOCUS       CA946732                 515 bp    mRNA    linear   EST 31-DEC-2002
DEFINITION  is09h12.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6364654 3',
            mRNA sequence.
ACCESSION   CA946732
VERSION     CA946732.1  GI:27439609
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 451.
FEATURES             Location/Qualifiers
     source          1..515
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6364654"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
BASE COUNT       94 a    154 c    114 g    153 t
ORIGIN

  Query Match             18.9%;  Score 188.8;  DB 14;  Length 515;
  Best Local Similarity   86.7%;  Pred. No. 9.8e+02;
  Matches  208;  Conservative    0;  Mismatches   32;  Indels    0;  Gaps    0;

Qy          6 AAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTTTGGGA 65
Db        297 AAAAATACAAAAATTAGGCTGGGCGCAGTGGCCCAGGCCTGTAATCCCAGCACTTTGGGA 238

Qy         66 GGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCG 125
Db        237 GGCCGAGGTGGGTGGATCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTG 178

Qy        126 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAAT 185
Db        177 AAACCCCATCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGTGTGCGCCTGTAAT 118

Qy        186 CCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db        117 CCCAGCTACTTGGGAGGCTGAGACAGGAGAATCGCTTGAACCCGGGAGGCAGAGGTTGCA 58


RESULT 15
AI339725/c
LOCUS       AI339725                 446 bp    mRNA    linear   EST 13-FEB-1999
DEFINITION  qg42e06.x1 Soares_NhHMPu_S1 Homo sapiens cDNA clone IMAGE:1935206
            3' similar to contains Alu repetitive element;contains element PTR5
            repetitive element ;, mRNA sequence.
ACCESSION   AI339725
VERSION     AI339725.1  GI:4076652
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 446)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 2084   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 446.
FEATURES             Location/Qualifiers
     source          1..446
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1935206"
                     /tissue_type="Pooled human melanocyte, fetal heart, and
                     pregnant uterus"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NhHMPu_S1"
                     /note="Organ: mixed (see below); Vector: pT7T3D-Pac
                     (Pharmacia) with a modified polylinker; Site_1: Not I;
                     Site_2: Eco RI; Equal amounts of plasmid DNA from three
                     normalized libraries (melanocyte 2NbHM, pregnant uterus
                     NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
                     were made in vitro. Following HAP purification, this DNA
                     was used as tracer in a subtractive hybridization
                     reaction. The driver was PCR-amplified cDNAs from pools of
                     5,000 clones made from the same 3 libraries. The pools
                     consisted of I.M.A.G.E. clones 260232-265223,
                     340488-345479, and 484488-489479."
BASE COUNT      120 a    117 c     99 g    110 t
ORIGIN

  Query Match             18.7%;  Score 187.4;  DB 9;  Length 446;
  Best Local Similarity   85.3%;  Pred. No. 1.2e+03;
```

```
Matches  209;  Conservative    0;  Mismatches   36;  Indels    0;  Gaps    0;

Qy          1 AATATAAAAATGCATACTTTAGGCTGGGCAAAGTGGCTCACGTCTGTAATCCCAGCACTT 60

Db        300 AATTTATTAAGAATTAATCTCGGCTGGGCGCGGTGGCTCACGCCTGTAATCCCAGCACTT 241

Qy         61 TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 120

Db        240 TGGGAAGCCAAGGTGGGCGGATCACTTCAGGTTAGGAGTTGGAGACCAGCCTGGCCAACA 181

Qy        121 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCT 180

Db        180 TGGCGAAACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCACTTGCCT 121

Qy        181 GTAATCCCAGCTACTCGGGAGGCTGAGACAGAAGAATTGCTTGAACCTTGGAGGGGGAGG 240

Db        120 GTAATCCCAGCTACTCGGGAGGCTGAGGCACGAGAATCGATTAAACCTGGGAGGCGGAGA 61

Qy        241 TTGCA 245
               |||
Db         60 TTGCA 56

Search completed: October 24, 2003, 22:51:13
Job time : 2095.5 secs
```

GenCore version 5.1.6

GM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1384.09 Seconds
(without alignments)
14809.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_10000_10500
Perfect score: 501
Sequence: 1 ggttcccgtcggtctgggg.....gggctgggggaggggcggc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454613386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgc_hum:*
40: em_htgc_mus:*
41: em_htgc_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		501	100.0	101584	9	CNS01DS5	AL121655 BAC seque
2		501	100.0	110000	6	AX093471	AX093471 Sequence
3		501	100.0	110000	9	HSA246003	AJ246003 Homo sapi
4		501	100.0	155943	9	AC012364	AC012364 Homo sapi
5		472	94.2	3263	6	AX093472	AX093472 Sequence
6		472	94.2	3263	9	HSA246001	AJ246001 Homo sapi
7		472	94.2	5120	9	AB029006	AB029006 Homo sapi
8		442	88.2	185281	2	AC011232	AC011232 Homo sapi
9		348.8	69.6	2152	6	AX093576	AX093576 Sequence
10		348.2	69.5	969	9	HSA325510	AJ325510 Homo sapi
11		346	69.1	2116	10	BC046286	BC046286 Mus muscu
12		274	54.7	71793	2	AC101733	AC101733 Mus muscu
13		247.4	49.4	721	9	HSA342191	AJ342191 Homo sapi
14		213.4	42.6	712	9	HSA342199	AJ342199 Homo sapi
15		165.8	33.1	697	9	HSA342196	AJ342196 Homo sapi
16		73	14.6	1689	6	AX093542	AX093542 Sequence
17		73	14.6	1689	10	MMU246002	AJ246002 Mus muscu
c	18	64	12.8	125020	9	AF429315	AF429315 Homo sapi
	19	59.4	11.9	161822	2	AC144461	AC144461 Mus muscu
c	20	58.6	11.7	133575	8	CNS0BC9N	AL732648 Oryza sat
	21	58.2	11.6	37159	2	AC100259	AC100259 Mus muscu
	22	57.2	11.4	65942	2	AC036124	AC036124 Homo sapi
	23	57	11.4	200542	9	CNS01RG3	AL157736 Human chr
	24	56.6	11.3	7218	6	I66494	I66494 Sequence 14
c	25	55.8	11.1	135119	2	AC011578	AC011578 Homo sapi
c	26	55.4	11.1	154729	9	AC018730	AC018730 Homo sapi
c	27	54.8	10.9	131602	2	AC034263	AC034263 Homo sapi
	28	54.4	10.9	80133	2	AC021347	AC021347 Homo sapi
c	29	54.2	10.8	63082	2	AC022663	AC022663 Homo sapi
c	30	53.8	10.7	159028	2	AC142003	AC142003 Pattus no
c	31	53.4	10.7	300695	2	AC079431	AC079431 Mus muscu
	32	53.2	10.6	87283	2	AC022856	AC022856 Homo sapi
	33	53.2	10.6	115402	2	AC027283	AC027283 Homo sapi
c	34	53	10.6	159851	2	AP003571	AP003571 Oryza sat
	35	53	10.6	171574	2	AC012300	AC012300 Homo sapi
	36	53	10.6	219952	2	AC084804	AC084804 Mus muscu
c	37	52.8	10.5	208220	9	AC009107	AC009107 Homo sapi
c	38	52.6	10.5	40343	2	AC099831	AC099831 Homo sapi
c	39	52.4	10.5	2837	9	HSCGG16P5	AJ001218 Homo sapi
	40	52.4	10.5	139087	2	AP004797	AP004797 Oryza sat
c	41	52.4	10.5	141534	2	AP003685	AP003685 Oryza sat
c	42	52.4	10.5	155419	2	AP005840	AP005840 Oryza sat
c	43	52.4	10.5	180865	2	AP003766	AP003766 Oryza sat
c	44	52.2	10.4	64211	2	AC091110	AC091110 Homo sapi
c	45	52.2	10.4	133790	2	AC119292	AC119292 Oryza sat

ALIGNMENTS

RESULT 1
CNS01DS5
LOCUS CNS01DS5 101584 bp DNA linear PRI 18-APR-2002
DEFINITION BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 336P14 of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION AL121655
VERSION AL121655.1 GI:6002388
KEYWORDS SPG4 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101584)
AUTHORS Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D., Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,

```
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'Homme,J.F., Brieu,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a novel AAA protein, is altered in the most frequent form
            of autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. (1999) In press
REFERENCE   2  (bases 1 to 101584)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1..101584
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="336P14"
                     /clone_lib="CITB_978_SKB"
BASE COUNT    29860 a   22110 c   22047 g   27567 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 9;  Length 101584;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-74;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      69973 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 70032

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70033 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 70092

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70093 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 70152

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70153 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 70212

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70213 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 70272

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70273 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 70332

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGGGCGTCCGAGTCTTCCACAAA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70333 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 70392

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70393 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 70452

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
              |||||||||||||||||||||
Db      70453 GGGGCTGGGGGAGGGGGCGGC 70473


RESULT 2
AX093471
LOCUS       AX093471               110000 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 1 from Patent WO0118198.
ACCESSION   AX093471
VERSION     AX093471.1  GI:13509911
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 1 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     exon            9932..10471
     intron          10472..33718
     exon            33719..33805
     intron          33806..35748
     exon            35749..35832
     intron          35833..45022
     exon            45023..45118
     intron          45119..60863
     exon            60864..61051
     intron          61052..61927
     exon            61928..62061
     intron          62062..62344
     exon            62345..62438
     intron          62439..73173
     exon            73174..73248
     intron          73249..74633
     exon            74634..74705
     intron          74706..82788
     exon            82789..82864
     intron          82865..83102
     exon            83103..83194
     intron          83195..83334
     exon            83335..83414
     intron          83415..88129
     exon            88130..88172
     intron          88173..89561
     exon            89562..89641
     intron          89642..91162
     exon            91163..91233
     intron          91234..93443
     exon            93444..93484
     intron          93485..100599
     exon            100600..102009
BASE COUNT    30622 a   21640 c   22817 g   34921 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 6;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 1.4e-74;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10000 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 10059

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10060 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 10119

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10120 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 10179

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10180 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 10239

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10240 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 10299
```

```
Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10300 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 10359

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10360 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 10419

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10420 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 10479

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
              |||||||||||||||||||||
Db      10480 GGGGCTGGGGGAGGGGGCGGC 10500


RESULT 3
HSA246003
LOCUS       HSA246003             110000 bp    DNA     linear   PRI 08-MAR-2000
DEFINITION  Homo sapiens Spast gene for spastin protein.
ACCESSION   AJ246003
VERSION     AJ246003.1  GI:6273492
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20035425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 110000)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The
            sequence is the result of the assembly of 2 BAC clones: R-336p14
            and
            563N4, respectively from RPCI-11 and CITB_978_SKB library.
FEATURES             Location/Qualifiers
     source          1. .110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
                     /clone="R-336P14"
                     /clone_lib="RPCI-11"
                     /note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
                     , BAC 563N4 contains Spast gene: exon 4 to exon 17"
     gene            9932. .102009
                     /gene="Spast"
     mRNA            9932. .102009
                     /gene="Spast"
                     /product="spastin protein"
     exon            9932. .10471
                     /gene="Spast"
                     /number=1
     5'UTR           9932. .10056
                     /gene="Spast"
     CDS             join(10057. .10471,33719. .33805,35749. .35832,45023. .45118,
                     60864. .61051,61928. .62061,62345. .62438,73174. .73248,
                     74634. .74705,82769. .82864,83103. .83194,83335. .83414,
                     88130. .88172,89562. .89641,91163. .91233,93444. .93484,
                     100600. .100722)
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60208.1"
                     /db_xref="GI:6273493"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     intron          10472. .33718
                     /gene="Spast"
                     /number=1
     exon            33719. .33805
                     /gene="Spast"
                     /number=2
     intron          33806. .35748
                     /gene="Spast"
                     /number=2
     exon            35749. .35832
                     /gene="Spast"
                     /number=3
     intron          35833. .45022
                     /gene="Spast"
                     /number=3
     exon            45023. .45118
                     /gene="Spast"
                     /number=4
     intron          45119. .60863
                     /gene="Spast"
                     /number=4
     exon            60864. .61051
                     /gene="Spast"
                     /number=5
     intron          61052. .61927
                     /gene="Spast"
                     /number=5
     exon            61928. .62061
                     /gene="Spast"
                     /number=6
     intron          62062. .62344
                     /gene="Spast"
                     /number=6
     exon            62345. .62438
                     /gene="Spast"
                     /number=7
     intron          62439. .73173
                     /gene="Spast"
                     /number=7
     exon            73174. .73248
                     /gene="Spast"
                     /number=8
     intron          73249. .74633
                     /gene="Spast"
```

```
                     /number=8
     exon            74634. .74705
                     /gene="Spast"
                     /number=9
     intron          74706. .82782
                     /gene="Spast"
                     /number=9
     exon            82783. .82864
                     /gene="Spast"
                     /number=10
     intron          82865. .83102
                     /gene="Spast"
                     /number=10
     exon            83103. .83194
                     /gene="Spast"
                     /number=11
     intron          83195. .83334
                     /gene="Spast"
                     /number=11
     exon            83335. .83414
                     /gene="Spast"
                     /number=12
     intron          83415. .88129
                     /gene="Spast"
                     /number=12
     exon            88130. .88172
                     /gene="Spast"
                     /number=13
     intron          88173. .89561
                     /gene="Spast"
                     /number=13
     exon            89562. .89641
                     /gene="Spast"
                     /number=14
     intron          89642. .91162
                     /gene="Spast"
                     /number=14
     exon            91163. .91233
                     /gene="Spast"
                     /number=15
     intron          91234. .93443
                     /gene="Spast"
                     /number=15
     exon            93444. .93484
                     /gene="Spast"
                     /number=16
     intron          93485. .100599
                     /gene="Spast"
                     /number=16
     exon            100600. .102009
                     /gene="Spast"
                     /number=17
     3'UTR           100721. .102009
                     /gene="Spast"
     polyA_signal    101973. .101978
                     /gene="Spast"
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match             100.0%;  Score 501;  DB 9;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-74;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10000 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 10059

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              |||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10060 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 10119

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              ||||||| ||||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Db      10120 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 10179

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              ||| |||| ||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10180 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 10239

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10240 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 10299

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10300 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 10359

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
              |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10360 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 10419

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
              |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db      10420 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 10479

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
              ||||||||||| |||  ||
Db      10480 GGGGCTGGGGGAGGGGGCGGC 10500


RESULT 4
AC012364
LOCUS       AC012364               155943 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-444D15 from 2, complete sequence.
ACCESSION   AC012364
VERSION     AC012364.7  GI:14589736
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155943)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 155943)
  AUTHORS   Ali,J. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-444D15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 155943)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:11602953.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0444D15
```

```
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 52:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-431F19; the clone sequenced
            to the right is RP11-78B13.  Actual start of this clone is at base
            position 1 of RP11-444D15; actual end is at base position 155943 of
            RP11-444D15.

            Data from AC011232 and AC010981 was used to finish this clone,
            AC012364.  The sequence fidelity between bases 112626 to 112643 can
            not be guaranteed due to an unresolved homopolymeric run. The
            sequence between 112648 to 112663 is single stranded.
FEATURES             Location/Qualifiers
     source          1..155943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-444D15"
                     /clone_lib="RPCI-11"
     repeat_region   310..617
                     /rpt_family="Alu"
     repeat_region   593..617
                     /rpt_family="AT_rich"
     repeat_region   728..748
                     /rpt_family="AT_rich"
     repeat_region   732..1026
                     /rpt_family="Alu"
     repeat_region   1169..1240
                     /rpt_family="(TATATG)n"
     repeat_region   1610..2503
                     /rpt_family="L2"
     repeat_region   2504..2776
                     /rpt_family="Alu"
     repeat_region   2777..3498
                     /rpt_family="L2"
     repeat_region   3590..3742
                     /rpt_family="MIR"
     repeat_region   4452..4765
                     /rpt_family="Alu"
     misc_feature    4550..4975
                     /note="match to EST AA632695 (NID:g2556109) np83b02.s1"
     misc_feature    4661..4692
                     /note="match to EST BG571163 (NID:g13578829)"
     repeat_region   4970..5239
                     /rpt_family="Alu"
     repeat_region   5211..5238
                     /rpt_family="(CAAAAA)n"
     misc_feature    5234..5328
                     /note="match to EST BG571163 (NID:g13578829)"
     misc_feature    5242..5328
                     /note="match to EST BG757662 (NID:g14068315)"
     misc_feature    5247..5328
                     /note="match to EST AU121379 (NID:g10936614)"
     misc_feature    5247..5328
                     /note="match to EST BF205430 (NID:g11099016)"
     misc_feature    5247..5328
                     /note="match to EST BG033287 (NID:g12425431)"
     misc_feature    5247..5328
                     /note="similar to Mus musculus EST AA619874 (NID:g2523750)
                     v159g02.s1"
     misc_feature    5247..5328
                     /note="similar to Mus musculus EST AW120623
                     (NID:g6095956)"
     repeat_region   5740..6346
                     /rpt_family="L2"
     repeat_region   6347..6438
                     /rpt_family="AcHobo"
     repeat_region   6453..6777
                     /rpt_family="L2"
     repeat_region   6938..7406
                     /rpt_family="L1"
     repeat_region   7415..8075
                     /rpt_family="ERV1"
     misc_feature    7900..8421
                     /note="match to EST AW994939 (NID:g8255173)"
     repeat_region   8087..8245
                     /rpt_family="L1"
     repeat_region   8447..8524
                     /rpt_family="MER2_type"
     repeat_region   8525..8799
                     /rpt_family="Alu"
     repeat_region   8800..9067
                     /rpt_family="MER2_type"
     repeat_region   9917..10069
                     /rpt_family="MER1_type"
     misc_feature    10293..10786
                     /note="match to EST N54249 (NID:g1195415) yv65c12.s1"
     repeat_region   10566..10753
                     /rpt_family="Alu"
     repeat_region   10820..11873
                     /rpt_family="L1"
     repeat_region   11588..11607
                     /rpt_family="(A)n"
     repeat_region   11863..11888
                     /rpt_family="AT_rich"
     repeat_region   11887..12044
                     /rpt_family="Tc2"
     repeat_region   12045..12345
                     /rpt_family="Alu"
     repeat_region   12382..12694
                     /rpt_family="Alu"
     repeat_region   12666..12727
                     /rpt_family="A-rich"
     repeat_region   12743..13296
                     /rpt_family="Tc2"
     repeat_region   13463..13492
                     /rpt_family="(T)n"
     misc_feature    13463..13478
                     /note="similar to Homo sapiens EST BG110835
                     (NID:g12604341)"
     misc_feature    13820..14405
                     /note="match to EST BG434159 (NID:g13340653)"
     repeat_region   14067..14097
                     /rpt_family="AT_rich"
     repeat_region   14793..15032
```

```
                     /rpt_family="L1"
     repeat_region   15033. .15330
                     /rpt_family="Alu"
     repeat_region   15331. .15399
                     /rpt_family="L2"
     repeat_region   15400. .15700
                     /rpt_family="Alu"
     misc_feature    15405. .15410
                     /note="match to EST BG576844 (NID:g13584497)"
     repeat_region   15701. .15745

  Query Match             100.0%;  Score 501;  DB 9;  Length 155843;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-74;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              |||||| |||| |||| |||| |||| |||| ||| ||| || || |
Db     125733 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 125792

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              |||||| .|||| ||| ||| |||| ||| ||| ||| |||| |||| |||| ||||
Db     125793 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 125852

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125853 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 125912

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              |||||||||||.|||||||||||||||||||||||||||||||||||||||||||||||
Db     125913 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 125972

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     125973 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 126032

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126033 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 126092

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126093 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 126152

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     126153 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 126212

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
              ||||||||||||||| ||||
Db     126213 GGGGCTGGGGGAGGGGGCGGC 126233


RESULT 5
AX093472
LOCUS       AX093472               3263 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 2 from Patent WO0118198.
ACCESSION   AX093472
VERSION     AX093472.1  GI:13509912
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 2 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      956 a    664 c    703 g    940 t
ORIGIN

  Query Match             94.2%;  Score 472;  DB 6;  Length 3263;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-69;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              | |||.|||| .|| |||| |||| |||| |||| |||| |||| |||| |||| |
Db         69 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 128

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              || ||| |.| |||.|.||| ||| .|||.|||| | |||| |||| |||| ||||
Db        129 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 188

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              |||.|||| ||| ||| |||| |||| .|||| .|||| |||| |||| |||| ||||
Db        189 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 248

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              |||| |||| |||| ||| |||| |||| |||| |||| |||| |||| |||| ||||
Db        249 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 308

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              ||| |||| ||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db        309 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 368

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              |||||||||||.|||||| |||| |||| |||| |||| |||| |||| |||| ||||
Db        369 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 428

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
              | |||.|||| |||.|||| |||||.|||| |||| |||| |||| |||| |||| |
Db        429 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 488

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
              ||| ||| .|||||||||||.|||| |||| |||| |||| |||| |||| |
Db        489 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 540


RESULT 6
HSA246001
LOCUS       HSA246001              3263 bp    mRNA    linear   PRI 10-MAR-2000
DEFINITION  Homo sapiens mRNA for spastin protein (Spast gene).
ACCESSION   AJ246001
VERSION     AJ246001.1  GI:6273490
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
```

```
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 3263)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr.
FEATURES             Location/Qualifiers
     source          1. .3263
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
     gene            1. .3263
                     /gene="Spast"
     CDS             136. .1976
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60141.1"
                     /db_xref="GI:6273491"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLESGAVPKRKDPLTHTSNSLPRSK
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     polyA_signal    3227. .3232
                     /gene="Spast"
BASE COUNT      956 a    664 c    703 g    940 t
ORIGIN

  Query Match             94.2%;  Score 472;  DB 9;  Length 3263;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-69;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         69 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 128

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        129 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 188

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        189 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 248

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 308

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 368

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 428

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 GCCTCGGCCCCGGCGCCGGTGCCGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 488

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 540


RESULT 7
AB029006
LOCUS       AB029006               5120 bp    mRNA    linear   PRI 04-AUG-1999
DEFINITION  Homo sapiens mRNA for KIAA1083 protein, complete cds.
ACCESSION   AB029006
VERSION     AB029006.1  GI:5689502
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Kikuno,R., Nagase,T., Ishikawa,K., Hirosawa,M., Miyajima,N.,
            Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
  TITLE     Prediction of the coding sequences of unidentified human genes.
            XIV. The complete sequences of 100 new cDNA clones from brain which
            code for large proteins in vitro
  JOURNAL   DNA Res. 6 (3), 197-205 (1999)
  MEDLINE   99397452
   PUBMED   10470851
REFERENCE   2  (bases 1 to 5120)
  AUTHORS   Ohara,O., Nagase,T. and Kikuno,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute,
            Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba
            292-0812, Japan (E-mail:odnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
            Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source          1. .5120
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="hj07423"
                     /tissue_type="brain"
                     /clone_lib="pBluescriptII SK plus"
     gene            1. .5120
                     /gene="KIAA1083"
     CDS             222. .1976
                     /gene="KIAA1083"
                     /codon_start=1
                     /product="KIAA1083 protein"
                     /protein_id="BAA83035.1"
                     /db_xref="GI:5689503"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLESGAVPKRKDPLTHTSNSLPRSK
                     TVMKTGSAGLSGHHRAPSYSGLSMVSGVKQGSGPAPTTHKGTPKTNRTNKPSTPTTAT
                     RKKKDLKNFRNVDSNLANLIMNEIVDNGTAVKFDDIAGQDLAKQALQEIVILPSLRPE
                     LFTGLRAPARGLLLFGPPGNGKTMLAKAVAAESNATFFNISAASLTSKYVGEGEKLVR
                     ALFAVARELQPSIIFIDEVDSLLCERREGEHDASRRLKTEFLIEFDGVQSAGDDRVLV
                     MGATNRPQELDEAVLRRFIKRVYVSLPNEETRLLLLKNLLCKQGSPLTQKELAQLARM
                     TDGYSGSDLTALAKDAALGPIRELKPEQVKNMSASEMRNIRLSDFTESLKKIKRSVSP
                     QTLEAYIRWNKDFGDTTV"
BASE COUNT     1532 a    943 c   1077 g   1568 t
ORIGIN

  Query Match             94.2%;  Score 472;  DB 9;  Length 5120;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-69;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        165 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 224

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        225 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 284
```

```
Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGGCGGGCCCCTCCG 180
Db        285 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGGCGGGCCCCTCCG 344

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db        345 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 404

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
Db        405 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 464

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
Db        465 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 524

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
Db        525 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 584

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
Db        585 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 636


RESULT 8
AC011232
LOCUS       AC011232              185281 bp    DNA     linear   HTG 10-MAR-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC011232
VERSION     AC011232.7  GI:13270720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 10, 2001 this sequence version replaced gi:9799811.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information ---  --------
            Center project name: H_NH0078E13
            -------------- Summary Statistics  ----- - ------
            Sequencing vector: M13; 57%
            Sequencing vector: plasmid; 40%
            Chemistry: Dye-primer ET; 48% of reads
            Chemistry: Dye-terminator Big Dye; 50% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182798 bases at least Q40
            Consensus quality: 183538 bases at least Q30
            Consensus quality: 184045 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 184681; sum-of-contigs
            Quality coverage: 7.07 in Q20 bases; agarose-fp
            Quality coverage: 7.33 in Q20 bases; sum-of-contigs
            ------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1985: contig of 1985 bp in length
            *    1986     2085: gap of unknown length
            *    2086     8183: contig of 6098 bp in length
            *    8184     8283: gap of unknown length
            *    8284    22741: contig of 14458 bp in length
            *   22742    22841: gap of unknown length
            *   22842    41446: contig of 18605 bp in length
            *   41447    41546: gap of unknown length
            *   41547    69182: contig of 27636 bp in length
            *   69183    69282: gap of unknown length
            *   69283   112131: contig of 42849 bp in length
            *  112132   112231: gap of unknown length
            *  112232   185281: contig of 73050 bp in length.
FEATURES             Location/Qualifiers
     source          1..185281
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-78E13"
     misc_feature    1..1985
                     /note="assembly_name:Contig11"
     misc_feature    2086..8183
                     /note="assembly_name:Contig12"
     misc_feature    8284..22741
                     /note="assembly_name:Contig13"
     misc_feature    22842..41446
                     /note="assembly_name:Contig14"
     misc_feature    41547..69182
                     /note="assembly_name:Contig15"
     misc_feature    69283..112131
                     /note="assembly_name:Contig16"
     misc_feature    112232..185281
                     /note="assembly_name:Contig17"
BASE COUNT    52719 a   36957 c   37187 g   57816 t     602 others
ORIGIN


  Query Match             88.2%;  Score 442;  DB 2;  Length 185281;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-65;
  Matches  442;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          60 GAATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCC 119
Db       22842 GAATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCC 22901

Qy         120 TCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGGCGGGCCCCTCC 179
Db       22902 TCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGGCGGGCCCCTCC 22961

Qy         180 GCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTT 239
Db       22962 GCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTT 23021

Qy         240 CGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCG 299
Db       23022 CGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCG 23081

Qy         300 CTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTC 359
Db       23082 CTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTC 23141

Qy         360 GGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAA 419
Db       23142 GGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAA 23201

Qy         420 ACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 479
Db       23202 ACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 23261
```

```
Qy        480 GGGGCTGGGGAGGGGGCGGC 501

Db      23262 GGGGCTGGGGAGGGGGCGGC 23283


RESULT 9
AX093576
LOCUS       AX093576                2152 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 106 from Patent WO0118198.
ACCESSION   AX093576
VERSION     AX093576.1  GI:13510014
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 106 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..2152
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /note="ADN Spg4 complet"
BASE COUNT      561 a    552 c    583 g    456 t
ORIGIN

  Query Match             69.6%;  Score 348.8;  DB 6;  Length 2152;
  Best Local Similarity   86.7%;  Pred. No. 7.3e-49;
  Matches  409;  Conservative    0;  Mismatches   57;  Indels    6;  Gaps    2;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60

Db         79 GGTCGCCGTCGGTCTGCGGGAAGCGGGTTATGGCGGCGGCGGCAGTGGGAGCTGTGAATG 138

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120

Db        139 AGTTCTCCGGCCGGACGACGGAAGAAGAAAGGCTCGGGCGGCGCGAG---CCCGGCGCCC 195

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180

Db        196 GCCAGGCCTCCGCCCCCGCCGCGGTCCCCGCCCCT---GCCGCCGGCCCGGCCCCTGCG 252

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240

Db        253 GCCGGCTCGCCGCCTAAGCGGAACCTGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTC 312

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300

Db        313 GCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGC 372

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360

Db        373 TTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCG 432

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420

Db        433 CCCTCGCCCCCAGCGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCACAAG 492

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472

Db        493 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAG 544


RESULT 10
HSA325510
LOCUS       HSA325510                969 bp    DNA     linear   PRI 16-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            NB6-679R.
ACCESSION   AJ325510
VERSION     AJ325510.1  GI:15869904
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 969)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2  (bases 1 to 969)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..969
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="NB6-679R"
BASE COUNT      159 a    281 c    324 g    205 t
ORIGIN

  Query Match             69.5%;  Score 348.2;  DB 9;  Length 969;
  Best Local Similarity   92.8%;  Pred. No. 1.1e 48;
  Matches  376;  Conservative    0;  Mismatches   28;  Indels    1;  Gaps    1;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60

Db        220 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGCGAATG 279

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120

Db        280 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 339

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180

Db        340 CCCAGGCCTTCGACCTCTTGCCTGGTCCCCGCCCCTCCCGACGGCGGGCCGGCCCCTGCG 399

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240

Db        400 TCCGAGTCGTCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 459

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300

Db        460 GCGCTGCTGCGTTTGGTCGCCTTGCACATGGGGCTCCTCTTTGTGTGGCTGTGCCAGCGC 519

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360

Db        520 TTGTTCCGCGCCCTCATGGCACCCAAGAGGAGCTGCGGAGCCGGGCCAGCAACTGCCTTG 579

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTC 405

Db        580 GCCTCGGGCCTCGCGCCAGTTCCGGGCGG-GAGGCCGAACGCGTC 623


RESULT 11
BC046286
LOCUS       BC046286                2116 bp    mRNA    linear   ROD 10-FEB-2003
DEFINITION  Mus musculus, spastic paraplegia 4 homolog (human), clone MGC:54786
            IMAGE:6441742, mRNA, complete cds.
ACCESSION   BC046286
VERSION     BC046286.1  GI:28279481
```

```
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2116)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
            Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 100 Row: i Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction, Similarity but not identity
            to protein.
FEATURES             Location/Qualifiers
     source          1..2116
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="MGC:54786 IMAGE:6441742"
                     /tissue_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month
                     old mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam2"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             51..1892
                     /codon_start=1
                     /product="spastic paraplegia 4 homolog (human)"
                     /protein_id="AAH46286.1"
                     /db_xref="GI:28279482"
                     /db_xref="LocusID:50850"
                     /translation="MSSPAGRRKKKGSGGASPAPARPPPPAAVPAPAAGPAPAAGSPP
                     KRNPSSFSSPLVVGFALLRLLACHLGLLFAWLCQRFSRALMAAKRSSGTAPAPASPSP
                     PAPGPGGEAESVRVFHKQAFEYISIALRIDEEEKGQKEQAVEWYKKGIEELEKGIAVI
                     VTGQGEQYERARRLQAKMMTNLVMAKDRLQLLEKLQPVLQFSKSQTDVYNESTNLTCR
                     NGHLQSESGAVPKRKDPLTHASNSLPRSKTVLKSGSAGLSGHHRAPSCSGLSMVSGAR
                     PGPGPAATTHKGTPKPNRTNKPSTPTTAVRKKKDLKNFRNVDSNLANLIMNEIVDNGT
                     AVKFDDIAGQELAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLAKAV
                     AAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCERREG
                     EHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSLPNE
                     ETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKPEQV
                     KNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
BASE COUNT      610 a    515 c    546 g    445 t
ORIGIN

  Query Match            69.1%;  Score 346;  DB 10;  Length 2116;
  Best Local Similarity  86.1%;  Pred. No. 2.1e-48;
  Matches  408;  Conservative    0;  Mismatches   60;  Indels    6;  Gaps    2;

Qy          8 GTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGAATTCTC 67

Db          1 GTCGGTCTGCGGGAAGCGGGTTATGGCGGCGGCGGCAGTGGGAGCTGTGAATGAGTTCTC 60

Qy         68 CGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTCCCAGGC 127

Db         61 CGGCCGGACGACGGAAGAAGAAAGGCTCGGGCGGCGCGAG---CCCGGCGCCCGCCAGGC 117

Qy        128 CTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGT 187

Db        118 CTCCGCCCCCCGCCGCGGTCCCCGCCCCT---GCCGCCGGCCCGGCCCCTGCGGCCGGCT 174

Qy        186 CGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGC 247

Db        175 CGCCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTCGCCCTGC 234

Qy        248 TGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCC 307

Db        235 TGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGCTTCTCCC 294

Qy        308 GCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGG 367

Db        295 GCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCGCCCTCGC 354

Qy        368 CCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCT 427

Db        355 CCCCAGCGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCT 414

Qy        428 TCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGG 481

Db        415 TCGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGACAGAAGG 468


RESULT 12
AC101733
LOCUS       AC101733               71793 bp    DNA     linear   HTG 23-NOV-2001
DEFINITION  Mus musculus clone RP23-417L6, LOW-PASS SEQUENCE SAMPLING.
ACCESSION   AC101733
VERSION     AC101733.1  GI:17060508
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 71793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-417L6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 71793)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Gnde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   All repeats were identified using RepeatMasker:
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          -------------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
          -------------- Project Information
              Center project name: LI7311
              Center clone name: 417_L_6
          --------------.
          * NOTE: This record contains 91 individual
          * sequencing reads that have not been assembled into
          * contigs. Runs of N are used to separate the reads
          * and the order in which they appear is completely
          * arbitrary. Low-pass sequence sampling is useful for
          * identifying clones that may be gene-rich and allows
          * overlap relationships among clones to be deduced.
          * However, it should not be assumed that this clone
          * will be sequenced to completion. In the event that
          * the record is updated, the accession number will
          * be preserved.
          *        1      690: contig of 690 bp in length
          *      691      790: gap of 100 bp
          *      791     1465: contig of 675 bp in length
          *     1466     1565: gap of 100 bp
          *     1566     2271: contig of 706 bp in length
          *     2272     2371: gap of 100 bp
          *     2372     3066: contig of 695 bp in length
          *     3067     3166: gap of 100 bp
          *     3167     3851: contig of 685 bp in length
          *     3852     3951: gap of 100 bp
          *     3952     4636: contig of 685 bp in length
          *     4637     4736: gap of 100 bp
          *     4737     5436: contig of 700 bp in length
          *     5437     5536: gap of 100 bp
          *     5537     6234: contig of 698 bp in length
          *     6235     6334: gap of 100 bp
          *     6335     7033: contig of 699 bp in length
          *     7034     7133: gap of 100 bp
          *     7134     7821: contig of 688 bp in length
          *     7822     7921: gap of 100 bp
          *     7922     8610: contig of 689 bp in length
          *     8611     8710: gap of 100 bp
          *     8711     9411: contig of 701 bp in length
          *     9412     9511: gap of 100 bp
          *     9512    10230: contig of 719 bp in length
          *    10231    10330: gap of 100 bp
          *    10331    11041: contig of 711 bp in length
          *    11042    11141: gap of 100 bp
          *    11142    11846: contig of 705 bp in length
          *    11847    11946: gap of 100 bp
          *    11947    12641: contig of 695 bp in length
          *    12642    12741: gap of 100 bp
          *    12742    13430: contig of 689 bp in length
          *    13431    13530: gap of 100 bp
          *    13531    14217: contig of 687 bp in length
          *    14218    14317: gap of 100 bp
          *    14318    15008: contig of 691 bp in length
          *    15009    15108: gap of 100 bp
          *    15109    15778: contig of 670 bp in length
          *    15779    15878: gap of 100 bp
          *    15879    16571: contig of 693 bp in length
          *    16572    16671: gap of 100 bp
          *    16672    17384: contig of 713 bp in length
          *    17385    17484: gap of 100 bp
          *    17485    18165: contig of 681 bp in length
          *    18166    18265: gap of 100 bp
          *    18266    18958: contig of 693 bp in length
          *    18959    19058: gap of 100 bp
          *    19059    19757: contig of 699 bp in length
          *    19758    19857: gap of 100 bp
          *    19858    20510: contig of 653 bp in length
          *    20511    20610: gap of 100 bp
          *    20611    21295: contig of 685 bp in length
          *    21296    21395: gap of 100 bp
          *    21396    22087: contig of 692 bp in length
          *    22088    22187: gap of 100 bp
          *    22188    22860: contig of 673 bp in length
          *    22861    22960: gap of 100 bp
          *    22961    23646: contig of 686 bp in length
          *    23647    23746: gap of 100 bp
          *    23747    24429: contig of 683 bp in length
          *    24430    24529: gap of 100 bp
          *    24530    25212: contig of 683 bp in length
          *    25213    25312: gap of 100 bp
          *    25313    25998: contig of 686 bp in length
          *    25999    26098: gap of 100 bp
          *    26099    26800: contig of 702 bp in length
          *    26801    26900: gap of 100 bp
          *    26901    27619: contig of 719 bp in length
          *    27620    27719: gap of 100 bp
          *    27720    28436: contig of 717 bp in length
          *    28437    28536: gap of 100 bp
          *    28537    29237  contig of 701 bp in length
          *    29238    29337: gap of 100 bp
          *    29338    30032: contig of 695 bp in length
          *    30033    30132: gap of 100 bp
          *    30133    30819: contig of 687 bp in length
          *    30820    30919: gap of 100 bp
          *    30920    31590: contig of 671 bp in length
          *    31591    31690: gap of 100 bp
          *    31691    32367: contig of 677 bp in length
          *    32368    32467: gap of 100 bp
          *    32468    33142: contig of 675 bp in length
          *    33143    33242: gap of 100 bp
          *    33243    33908: contig of 666 bp in length
          *    33909    34008: gap of 100 bp
          *    34009    34697: contig of 689 bp in length
          *    34698    34797: gap of 100 bp
          *    34798    35465: contig of 668 bp in length
          *    35466    35565: gap of 100 bp
          *    35566    36268: contig of 703 bp in length
          *    36269    36368: gap of 100 bp
          *    36369    37073: contig of 705 bp in length
          *    37074    37173: gap of 100 bp
          *    37174    37858: contig of 685 bp in length
          *    37859    37958: gap of 100 bp
          *    37959    38638: contig of 680 bp in length
          *    38639    38738: gap of 100 bp
          *    38739    39436: contig of 698 bp in length
          *    39437    39536: gap of 100 bp
          *    39537    40211: contig of 675 bp in length
          *    40212    40311: gap of 100 bp
          *    40312    41010: contig of 699 bp in length
          *    41011    41110: gap of 100 bp
          *    41111    41811: contig of 701 bp in length
          *    41812    41911: gap of 100 bp
          *    41912    42612: contig of 701 bp in length
          *    42613    42712: gap of 100 bp
          *    42713    43407: contig of 695 bp in length
          *    43408    43507: gap of 100 bp
          *    43508    44181: contig of 674 bp in length
          *    44182    44281: gap of 100 bp
          *    44282    44964: contig of 683 bp in length
          *    44965    45064: gap of 100 bp
          *    45065    45778: contig of 714 bp in length
          *    45779    45878: gap of 100 bp
          *    45879    46577: contig of 699 bp in length
          *    46578    46677: gap of 100 bp
          *    46678    47362: contig of 685 bp in length
          *    47363    47462: gap of 100 bp
```

```
        •   47463    48154: contig of 692 bp in length
        •   48155    48254: gap of 100 bp
        •   48255    48953: contig of 699 bp in length
        •   48954    49053: gap of 100 bp
        •   49054    49738: contig of 685 bp in length
        •   49739    49838: gap of 100 bp
        •   49839    50526: contig of 688 bp in length
        •   50527    50626: gap of 100 bp
        •   50627    51321: contig of 695 bp in length
        •   51322    51421: gap of 100 bp
        •   51422    52113: contig of 692 bp in length
        •   52114    52213: gap of 100 bp
        •   52214    52909: contig of 696 bp in length
        •   52910    53009: gap of 100 bp
        •   53010    53669: contig of 660 bp in length
        •   53670    53769: gap of 100 bp
        •   53770    54462: contig of 693 bp in length
        •   54463    54562: gap of 100 bp

  Query Match             54.7%;  Score 274;  DB 2;  Length 71793;
  Best Local Similarity   81.8%;  Pred. No. 1.1e-36;
  Matches  342;  Conservative    0;  Mismatches   70;  Indels    6;  Gaps    2;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
Db      57132 GGTCGCCGTCGGTCTGCGGGAAGCGGGTTATGGCGGCGGCGGTAGTGGGAGCTGTGAATG 57191

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
Db      57192 AGTTCTCCGGCCGGACGACGGAAGAAGAAAGGCTCGGGCGGCGCGAG---CCCGGCGCCC 57248

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
Db      57249 GCCAGGCCTCCGCCCCCCGCCGCGGTCCCCGCCCCT---GCCGCCGGCCCGGCCCCTGCG 57305

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db      57306 GCCGGCTCGCCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTC 57365

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
Db      57366 CCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGATCTGCCAGCGC 57425

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
Db      57426 TTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCGGGGATCGCATCGGCGCCGGCCTCA 57485

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCCGTCCGAGTCTTCCACA 418
Db      57486 CCCTCGCCCCCAGAGCCCAGACCGGGTGGTCAGGGCAAAAGCGTCCGCATCTTCCACA 57543

RESULT 13
HSA342191
LOCUS       HSA342191                721 bp    DNA     linear   PRI 19-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            HSJ-6AN6RS.
ACCESSION   AJ342191
VERSION     AJ342191.1  GI:15886608
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 721)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2  (bases 1 to 721)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..721
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="HSJ-6AN6RS"
BASE COUNT      115 a    215 c    260 g    128 t      3 others
ORIGIN

  Query Match             49.4%;  Score 247.4;  DB 9;  Length 721;
  Best Local Similarity   95.0%;  Pred. No. 7.4e-32;
  Matches  266;  Conservative    0;  Mismatches   12;  Indels    2;  Gaps    1;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
Db        220 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 279

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
Db        280 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 339

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
Db        340 TCCAGGCCTCCGCCCGCTTGCCTGGCCCCCGCCCCTCCCGCCGGCGGGCCGGCCCCTCCG 399

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db        400 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATNTTTCCTACCCGCTGTTTGTAGGTTTC 459

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCT 280
Db        460 GCGCTGTT--GTTTGGGTCCCTTCCACCTGGGGCTTCTTT 497

RESULT 14
HSA342199
LOCUS       HSA342199                712 bp    DNA     linear   PRI 18-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            HSJ-6AJ4RS.
ACCESSION   AJ342199
VERSION     AJ342199.1  GI:15886616
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 712)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2  (bases 1 to 712)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..712
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
                    /db_xref="taxon:9606"
                    /clone="HSJ-6AJ4RS"
BASE COUNT      112 a    217 c    253 g    130 t
ORIGIN

  Query Match             42.6%;  Score 213.4;  DB 9;  Length 712;
  Best Local Similarity   89.7%;  Pred. No. 3.5e-26;
  Matches  252;  Conservative    0;  Mismatches   26;  Indels    3;  Gaps    2;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60

Db        220 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 279

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120

Db        280 AATTCTCCGGGTGGACGAAGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 339

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180

Db        340 CCCATGCCTTCGCCCCCTTGCCTGGCCCCCGCCCCTGCCGGCGCCGGGCCGGCCACCTCG 399

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240

Db        400 CCCGAGTTGCCGGATAAGCGGAACCTGT-CTATTTCTCCTACCCCTTGTTTG--AGGCTC 456

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTT 281

Db        457 CCCCTTGTTCGGTTGGCTCCCTTCCACCTGGGGCTCTTTTT 497


RESULT 15
HSA342196
LOCUS       HSA342196               697 bp    DNA     linear   PRI 18-JUL-2002
DEFINITION  Homo sapiens genomic sequence surrounding NotI site, clone
            HSJ-6AI4RS.
ACCESSION   AJ342196
VERSION     AJ342196.1  GI:15886613
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 697)
  AUTHORS   Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
            Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
            Levitsky,V.G., Kolchanov,N.A., Protopopov,A.I., Kashuba,V.I.,
            Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovsky,E.R.
  TITLE     NotI flanking sequences: a tool for gene discovery and verification
            of the human genome
  JOURNAL   Nucleic Acids Res. 30 (14), 3163-3170 (2002)
  MEDLINE   22131767
   PUBMED   12136098
REFERENCE   2  (bases 1 to 697)
  AUTHORS   Zabarovsky,E.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
            Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
            Sweden
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="HSJ-6AI4RS"
BASE COUNT      112 a    209 c    239 g    136 t      1 others
ORIGIN

  Query Match             33.1%;  Score 165.8;  DB 9;  Length 697;
  Best Local Similarity   87.6%;  Pred. No. 3.1e-18;
  Matches  204;  Conservative    0;  Mismatches   27;  Indels    2;  Gaps    2;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCA-GTGAGAGCTGTGAAT 59

Db        220 GGTTACCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCATGTGAGAGCTGTGAAT 279

Qy         60 GAATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCC 119

Db        280 GAATTCTCCGGGTGGACGACGGAAGAAGAAAGGCTGCGGCGGCACCAGCAACCCGGTGCC 339

Qy        120 TCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCC 179

Db        340 TCCCAGGCCTGCGCCCCCTTGCCTGGCCCACGTCTCTCCCGTCGCCGGGCAGGTCACTCC 399

Qy        180 GCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTG 232

Db        400 GGCTGAGTCCCCG-ATTATCGGAACCTTTACTATTTTGCTACCCCTTTTTTAG 451

Search completed: October 24, 2003, 21:06:06
Job time : 1388.29 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 126.251 Seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US 09-830-902-1_COPY_10000_10500
Perfect score: 501
Sequence: 1 ggttcccgtcggtctgcggg.....gggctgggggaggggcggc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	110000	22	AAF84800	Nucleotide sequenc
	2	472	94.2	3263	22	AAF84801	Nucleotide sequenc
c	3	315.4	63.0	777	24	ABQ39732	Oligonucleotide fo
	4	315.4	63.0	777	24	ABQ39733	Oligonucleotide fo
	5	302.6	60.4	777	24	ABQ39730	Oligonucleotide fo
c	6	302.6	60.4	777	24	ABQ39731	Oligonucleotide fo
c	7	229.8	45.9	562	24	ABQ24124	Oligonucleotide fo
	8	229.8	45.9	562	24	ABQ24125	Oligonucleotide fo
	9	186.8	37.3	562	24	ABQ24122	Oligonucleotide fo
c	10	186.8	37.3	562	24	ABQ24123	Oligonucleotide fo
	11	73	14.6	1689	22	AAF84870	Nucleotide sequenc
	12	50.8	10.1	500	24	ABS71875	Human GTP-Rho bind
	13	49.6	9.9	2795	22	AAI71296	Streptomyces grise
	14	49.6	9.9	2795	22	AAF81954	Streptomyces grise
	15	49.4	9.9	1269	17	AAT33617	DxrA gene encoding
c	16	49.4	9.9	2870	21	AAZ31413	Complementary DNA
c	17	48.2	9.6	1127	21	AAA02477	Human colon cancer
	18	47.8	9.5	1700	14	AAQ45569	Sequence scyC and
	19	46.8	9.3	5760	24	ABS78661	M. echinospora DNA
	20	46.6	9.3	31168	22	ABA07295	Human pancreatic c
	21	46.6	9.3	31168	22	AAS42115	Genomic sequence #
	22	46.6	9.3	31168	22	AAS27804	DNA encoding novel
c	23	46.6	9.3	31168	22	AAK64760	Human immune/haema
	24	46.6	9.3	31168	22	AAK89940	Human digestive sy
	25	46.4	9.3	12001	16	AAQ76213	HSV L/ST region.
	26	46.2	9.2	2081	19	AAV01452	Modified dcxA gene
	27	46.2	9.2	3013	19	AAV01461	Plasmid pANT195 in
	28	46.2	9.2	3196	19	AAV01463	Streptomyces dauno
	29	46.2	9.2	3196	19	AAV01447	Daunomycin C-14 hy
	30	46	9.2	1504	25	ACA56678	Human signalling p
	31	46	9.2	1504	25	ABZ42657	Human dopamine rec
	32	46	9.2	1608	17	AAT27547	Recombinant human
	33	46	9.2	1610	15	AAQ57657	Allele D4.7 of the
	34	45.8	9.1	114955	20	AAX53491	Human adenosine A1
c	35	45.6	9.1	1503	22	AAH15281	Human cDNA sequenc
c	36	45.6	9.1	2607	24	ABS52437	Human cDNA for MDD
	37	45	9.0	3177	17	AAT16474	B virus gB glycopr
	38	45	9.0	3177	19	AAV33167	Simian herpesvirus
c	39	44.4	8.9	114955	20	AAX53491	Human adenosine A1
c	40	44.2	8.8	761	24	ABQ40806	Oligonucleotide fo
	41	44.2	8.8	761	24	ABQ40807	Oligonucleotide fo
c	42	44.2	8.8	1325	20	AAX90924	Epstein Barr Virus
	43	44.2	8.8	154746	24	AAD25519	Human herpesvirus
c	44	44.2	8.8	154746	24	AAD25519	Human herpesvirus
c	45	44	8.8	8438	15	AAQ73500	DNA encoding Pseud

ALIGNMENTS

```
RESULT 1
AAF84800
ID   AAF84800 standard; DNA; 110000 BP.
XX
AC   AAF84800;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of the human SPG4 gene.
XX
KW   Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9932..102009
FT                   /*tag= a
FT                   /note= "contains introns"
FT   exon            9932..10471
FT                   /*tag= b
FT                   /number= "1"
FT   intron          10472..33718
FT                   /*tag= c
FT                   /number= "1"
FT   exon            33719..33805
FT                   /*tag= d
FT                   /number= "2"
FT   intron          33806..35748
FT                   /*tag= e
FT                   /number= "2"
```

```
FT   exon            35749..35832
FT                   /*tag= f
FT                   /number= "3"
FT   intron          35833..45022
FT                   /*tag= g
FT                   /number= "3"
FT   exon            45023..45118
FT                   /*tag= h
FT                   /number= "4"
FT   intron          45119..60863
FT                   /*tag= i
FT                   /number= "4"
FT   exon            60864..61051
FT                   /*tag= j
FT                   /number= "5"
FT   intron          61052..61927
FT                   /*tag= k
FT                   /number= "5"
FT   exon            61928..62061
FT                   /*tag= l
FT                   /number= "6"
FT   intron          62062..62344
FT                   /*tag= m
FT                   /number= "6"
FT   exon            62345..62438
FT                   /*tag= n
FT                   /number= "7"
FT   intron          62439..73173
FT                   /*tag= o
FT                   /number= "7"
FT   exon            73174..73248
FT                   /*tag= p
FT                   /number= "8"
FT   intron          73249..74633
FT                   /*tag= q
FT                   /number= "8"
FT   exon            74634..74705
FT                   /*tag= r
FT                   /number= "9"
FT   intron          74706..82788
FT                   /*tag= s
FT                   /number= "9"
FT   exon            82789..82864
FT                   /*tag= t
FT                   /number= "10"
FT   intron          82865..83102
FT                   /*tag= u
FT                   /number= "10"
FT   exon            83103..83194
FT                   /*tag= v
FT                   /number= "11"
FT   intron          83195..83334
FT                   /*tag= w
FT                   /number= "11"
FT   exon            83335..83414
FT                   /*tag= x
FT                   /number= "12"
FT   intron          83415..88129
FT                   /*tag= y
FT                   /number= "12"
FT   exon            88130..88172
FT                   /*tag= z
FT                   /number= "13"
FT   intron          88173..89561
FT                   /*tag= aa
FT                   /number= "13"
FT   exon            89562..89641
FT                   /*tag= ab
FT                   /number= "14"
FT   intron          89642..91162
FT                   /*tag= ac
FT                   /number= "14"
FT   exon            91163..91233
FT                   /*tag= ad
FT                   /number= "15"
FT   intron          91234..93443
FT                   /*tag= ae
FT                   /number= "15"
FT   exon            93444..93484
FT                   /*tag= af
FT                   /number= "16"
FT   intron          93485..100599
FT                   /*tag= ag
FT                   /number= "16"
FT   exon            100600..102009
FT                   /*tag= ah
FT                   /number= "17"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening  -
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 other;

  Query Match             100.0%;  Score 501;  DB 22;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-92;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10000 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 10059

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10060 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 10119

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10120 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 10179

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10180 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 10239

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      10240 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 10299

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    10300 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 10359

Qy      361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420

Db    10360 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 10419

Qy      421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480

Db    10420 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 10479

Qy      481 GGGGCTGGGGGAGGGGGCGGC 501

Db    10480 GGGGCTGGGGGAGGGGGCGGC 10500
```

RESULT 2
AAF84801
ID   AAF84801 standard; cDNA; 3263 BP.
XX
AC   AAF84801;
XX
DT   09-JUL-2001 (first entry)
XX
DE   Nucleotide sequence of the human SPG4 polypeptide.
XX
KW   Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   126..1976
FT   /*tag= a
FT   /product= "spastin"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J, Hazan J;
XX
DR   WPI; 2001-283966/30.
DR   P-PSDB; AAB68137.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening -
XX
PS   Claim 2; Page 106-108; 145pp; French.
XX
CC   The present sequence encodes a human spastin polypeptide. Spastin is
CC   encoded by the SPG4 gene. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 3263 BP; 956 A; 664 C; 703 G; 940 T; 0 other;

```
  Query Match             94.2%;  Score 472;  DB 22;  Length 3263;
  Best Local Similarity   100.0%;  Pred. No. 5.2e-86;
  Matches  472;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60

Db       69 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 128

Qy       61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120

Db      129 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 188

Qy      121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180

Db      189 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 248

Qy      181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240

Db      249 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 308

Qy      241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300

Db      309 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 368

Qy      301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360

Db      369 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 428

Qy      361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420

Db      429 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 488

Qy      421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472

Db      489 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 540
```

RESULT 3
ABQ39732/c
ID   ABQ39732 standard; DNA; 777 BP.
XX
AC   ABQ39732;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 26323.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert

```
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 777 BP; 135 A; 79 C; 261 G; 302 T; 0 other;

  Query Match             63.0%;  Score 315.4;  DB 24;  Length 777;
  Best Local Similarity   78.8%;  Pred. No. 1.4e-54;
  Matches  376;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy          3 TTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGAA 62
Db        668 TTCCCGTCGATCTACGAAAAACGAATTATAACGACGACGACAATAAAAACTATAAATAAA 609

Qy         63 TTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTCC 122
Db        608 TTCTCCGAATAAACGAAAAAAAAAAAAAAACTCCGACGACGCCAACAACCCGATACCTCC 549

Qy        123 CAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCC 182
Db        548 CAAACCTCCGCCCCCTTACCTAACCCCCGCCCCTCCCGCCGCCGAACCGACCCCTCCGCC 489

Qy        183 CGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGC 242
Db        488 CGAATCGCCGCATAAACGAAACCTATACTATTTCTCCTACCCGCTATTTATAAACTTCGC 429

Qy        243 GCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTT 302
Db        428 GCTACTACGTTTAATCGCCTTCCACCTAAAACTCCTCTTCGTATAACTCTACCAACGCTT 369

Qy        303 CTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGC 362
Db        368 CTCCCGCGCCCTCATAACAACCAAAAAAAACTCCGAAACCGCGCCAACACCTACCTCGAC 309

Qy        363 CTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACA 422
Db        308 CTCGACCCCGACGCCGATACCGAACGACGAAACCGAACGCGTCCGAATCTTCCACAAACA 249

Qy        423 GGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 479
Db        248 AACCTTCGAATACATCTCCATTACCCTACGCATCGATAAAAATAAAAAAAATAACTA 192
```

```
RESULT 4
ABQ39733
ID   ABQ39733 standard; DNA; 777 BP.
XX
AC   ABQ39733;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 26324.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 777 BP; 302 A; 261 C; 79 G; 135 T; 0 other;

  Query Match             63.0%;  Score 315.4;  DB 24;  Length 777;
  Best Local Similarity   78.8%;  Pred. No. 1.4e-54;
  Matches  376;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy          3 TTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGAA 62
Db        110 TTCCCGTCGATCTACGAAAAACGAATTATAACGACGACGACAATAAAAACTATAAATAAA 169

Qy         63 TTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTCC 122
Db        170 TTCTCCGAATAAACGAAAAAAAAAAAAAAACTCCGACGACGCCAACAACCCGATACCTCC 229

Qy        123 CAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCC 182
Db        230 CAAACCTCCGCCCCCTTACCTAACCCCCGCCCCTCCCGCCGCCGAACCGACCCCTCCGCC 289

Qy        183 CGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGC 242
Db        290 CGAATCGCCGCATAAACGAAACCTATACTATTTCTCCTACCCGCTATTTATAAACTTCGC 349

Qy        243 GCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTT 302
Db        350 GCTACTACGTTTAATCGCCTTCCACCTAAAACTCCTCTTCGTATAACTCTACCAACGCTT 409

Qy        303 CTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGC 362
Db        410 CTCCCGCGCCCTCATAACAACCAAAAAAAACTCCGAAACCGCGCCAACACCTACCTCGAC 469
```

```
Qy        363 CTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACA 422
Db        473 CTCGACCCCGACGCCGATACCGAACGACGAAACCGAACGCGTCCGAATCTTCCACAAACA 529

Qy        423 GGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 479
Db        530 AACCTTCGAATACATCTCCATTACCCTACGCATCGATAAAAATAAAAAAAATAACTA 586
```

RESULT 5
ABQ39730
ID ABQ39730 standard; DNA; 777 BP.
XX
AC ABQ39730;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26321.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 777 BP; 106 A; 79 C; 275 G; 317 T; 0 other;

Query Match 60.4%; Score 302.6; DB 24; Length 777;
Best Local Similarity 75.2%; Pred. No. 5.3e-52;
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```
Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
Db        108 GGTTTTCGTCGGTTTGCGGGAGGCGGGTTATGGCGGCGGCGGTAGTGAGAGTTGTGAATG 167

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
Db        168 AATTTTTCGGGTGGACGAGGGAAGAAGAAAGGTTTCGGCGGCGTTAGTAATTCGGTGTTT 227

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
Db        228 TTTAGGTTTTCGTTTTTTTGTTTGGTTTTCGTTTTTTTCGTCGTCGGGTCGGTTTTTTCG 287

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db        288 TTCGAGTCGTCGTATAAGCGGAATTTGTATTATTTTTTTTATTCGTTGTTTGTAGGTTTC 347

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
Db        348 GCGTTGTTGCGTTTGGTCGTTTTTTATTTGGGGTTTTTTTTCGTGTGGTTTTGTTAGCGT 407

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
Db        408 TTTTTTCGCGTTTTTATGGTAGTTAAGAGGAGTTTCGGGGTCGCGTTAGTATTTGTTTCG 467

Qy        361 GCCTCGGCCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
Db        468 GTTTCGGTTTCGGCGTCGGTGTCGGGCGGCGAGGTCGAGCGCGTTCGAGTTTTTTATAAA 527

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
Db        528 TAGGTTTTCGAGTATATTTTTATTGTTTTGCGTATCGATGAGGATGAGAAAGGTAATTAG 587

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
Db        588 GGGGTTGGGGGAGGGGCGGC 608
```

RESULT 6
ABQ39731/C
ID ABQ39731 standard; DNA; 777 BP.
XX
AC ABQ39731;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26322.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -

```
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 777 BP; 317 A; 275 C; 79 G; 106 T; 0 other;

  Query Match             60.4%;  Score 302.6;  DB 24;  Length 777;
  Best Local Similarity   75.2%;  Pred. No. 5.3e-52;
  Matches  377;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
Db        670 GGTTTTCGTCGGTTTGCGGGAGGCGGGTTATGGCGGCGGCGGTAGTGAGAGTTGTGAATG 611

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
Db        610 AATTTTTCGGGTGGACGAGGGAAGAAGAAAGGTTTCGGCGGCGTTAGTAATTCGGTGTTT 551

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
Db        550 TTTAGGTTTTCGTTTTTTTGTTTGGTTTTCGTTTTTTTCGTCGTCGGGTCGGTTTTTTCG 491

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db        490 TTCGAGTCGTCGTATAAGCGGAATTTGTATTATTTTTTTTATTCGTTGTTTGTAGGTTTC 431

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
Db        430 GCGTTGTTGCGTTTGGTCGTTTTTTATTTGGGGTTTTTTTTCGTGTGGTTTTGTTAGCGT 371

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
Db        370 TTTTTTCGCGTTTTTATGGTAGTTAAGAGGAGTTTCGGGGTCGCGTTAGTATTTGTTTCG 311

Qy        361 GCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAA 420
Db        310 GTTTCGGTTTCGGCGTCGGTGTCGGGCGGCGAGGTCGAGCGCGTTCGAGTTTTTTATAAA 251

Qy        421 CAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAG 480
Db        250 TAGGTTTTCGAGTATATTTTTATTGTTTTGCGTATCGATGAGGATGAGAAAGGTAATTAG 191

Qy        481 GGGGCTGGGGGAGGGGGCGGC 501
Db        190 GGGGTTGGGGGAGGGGGCGGC 170

RESULT 7
ABQ24124/c
ID   ABQ24124 standard; DNA; 562 BP.
XX
AC   ABQ24124;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 10715.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 562 BP; 113 A; 53 C; 195 G; 201 T; 0 other;

  Query Match             45.9%;  Score 229.8;  DB 24;  Length 562;
  Best Local Similarity   80.3%;  Pred. No. 2.2e-37;
  Matches  282;  Conservative    0;  Mismatches   67;  Indels    2;  Gaps    1;

Qy        129 TCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTC 188
Db        562 TCCGCCCCCTTACCTAACCCCCGCCCCTCCCGCCGCCGAACCGACCCCTCCGCCCGAATC 503

Qy        189 GCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCT 248
Db        502 GCCGCATAAACGAAACCTATACTATTTCTCCTACCCGCTATTTATAAACTTCGCGCTACT 443

Qy        249 GCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCG 308
Db        442 ACGTTTAATCGCCTTCCACCTAAAACTCCTCTTCGTATAACTCTACCAACGCTTCTCCCG 383

Qy        309 CGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGC 368
```

```
Db        382 CGCCCTCATAACAACCAAAAAAAAACTCCGAAACCGCGCCAACACCTACCTCGACCTCGAC 323

Qy        369 CCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTT 428

Db        322 CC--GACGCCGATACCGACGACGAAACCGAACGCGTCCGAATCTTCCACAAACAAACCTT 265

Qy        429 CGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 479

Db        264 CGAATACATCTCCATTACCCTACGCATCGATAAAAATAAAAAAAATAACTA 214


RESULT 8
ABQ24125
ID   ABQ24125 standard; DNA; 562 BP.
XX
AC   ABQ24125;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 10716.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 562 BP; 201 A; 195 C; 53 G; 113 T; 0 other;

  Query Match             45.9%;  Score 229.8;  DB 24;  Length 562;
  Best Local Similarity   80.3%;  Pred. No. 2.2e-37;
  Matches  282;  Conservative    0;  Mismatches   67;  Indels    2;  Gaps    1;

Qy        129 TCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTC 188

Db          1 TCCGCCCCCTTACCTAACCCCCGCCCCTCCCGCCGCCGAACCGACCCCTCCGCCCGAATC 60

Qy        189 GCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCT 248

Db         61 GCCGCATAAACGAAACCTATACTATTTCTCCTACCCGCTATTTATAAACTTCGCGCTACT 120

Qy        249 GCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCG 308

Db        121 ACGTTTAATCGCCTTCCACCTAAAACTCCTCTTCGTATAACTCTACCAACGCTTCTCCCG 180

Qy        309 CGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCCGCGCCAGCACCTGCCTCGGCCTCGGC 368

Db        181 CGCCCTCATAACAACCAAAAAAAAACTCCGAAACCGCGCCAACACCTACCTCGACCTCGAC 240

Qy        369 CCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTT 428

Db        241 CC--GACGCCGATACCGACGACGAAACCGAACGCGTCCGAATCTTCCACAAACAAACCTT 298

Qy        429 CGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTA 479

Db        299 CGAATACATCTCCATTACCCTACGCATCGATAAAAATAAAAAAAATAACTA 349


RESULT 9
ABQ24122
ID   ABQ24122 standard; DNA; 562 BP.
XX
AC   ABQ24122;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 10713.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
```

CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 562 BP; 75 A; 53 C; 179 G; 255 T; 0 other;

Query Match 37.3%; Score 186.8; DB 24; Length 562;
Best Local Similarity 71.4%; Pred. No. 9.7e-29;
Matches 260; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

```
Qy        138 TTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAA 197
Db         10 TTGTTTGGTTTTCGTTTTTTTCGTCGTCGGGTCGGTTTTTTCGTTCGAGTCGTCGTATAA  69

Qy        198 GCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGT 257
Db         70 GCGGAATTTGTATTATTTTTTTTATTCGTTGTTTGTAGGTTTCGCGTTGTTGCGTTTGGT 129

Qy        258 CGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCAT 317
Db        130 CGTTTTTTATTTGGGGTTTTTTTTCGTGTGGTTTTGTTAGCGTTTTTTCGCGTTTTTAT 189

Qy        318 GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCC 377
Db        190 GGTAGTTAAGAGGAGTTTCGGGGTCGCGTTAGTATTTGTTTCGGTTTCGGTTCGGCGTCG 249

Qy        378 GGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACAT 437
Db        250 GTGTC--GGCGGCGAGGTCGAGCGCGTTCGAGTTTTTTATAAATAGGTTTTCGAGTATAT 307

Qy        438 CTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGGGGGCTGGGGGAGGGGG 497
Db        308 TTTTATTGTTTTGCGTATCGATGAGGATGAGAAAGGTAATTAGGGGGTTGGGGGAGGGGG 367

Qy        498 CGGC 501
Db        368 CGGC 371
```

RESULT 10
ABQ24123/c
ID ABQ24123 standard; DNA; 562 BP.
XX
AC ABQ24123;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 10714.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 562 BP; 255 A; 179 C; 53 G; 75 T; 0 other;

Query Match 37.3%; Score 186.8; DB 24; Length 562;
Best Local Similarity 71.4%; Pred. No. 9.7e-29;
Matches 260; Conservative 0; Mismatches 102; Indels 2; Gaps 1;

```
Qy        138 TTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAA 197
Db        553 TTGTTTGGTTTTCGTTTTTTTCGTCGTCGGGTCGGTTTTTTCGTTCGAGTCGTCGTATAA 494

Qy        198 GCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGT 257
Db        493 GCGGAATTTGTATTATTTTTTTTATTCGTTGTTTGTAGGTTTCGCGTTGTTGCGTTTGGT 434

Qy        258 CGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCAT 317
Db        433 CGTTTTTTATTTGGGGTTTTTTTTCGTGTGGTTTTGTTAGCGTTTTTTCGCGTTTTTAT 374

Qy        318 GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCC 377
Db        373 GGTAGTTAAGAGGAGTTTCGGGGTCGCGTTAGTATTTGTTTCGGTTTCGGTTCGGCGTCG 314

Qy        378 GGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACAT 437
Db        313 GTGTC--GGCGGCGAGGTCGAGCGCGTTCGAGTTTTTTATAAATAGGTTTTCGAGTATAT 256

Qy        438 CTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGGGGGCTGGGGGAGGGGG 497
Db        255 TTTTATTGTTTTGCGTATCGATGAGGATGAGAAAGGTAATTAGGGGGTTGGGGGAGGGGG 196

Qy        498 CGGC 501
Db        195 CGGC 192
```

RESULT 11
AAF84870
ID AAF84870 standard; cDNA; 1689 BP.

```
XX
AC   AAF84870;
XX
DT   09-JUL-2001  (first entry)
XX
DE   Nucleotide sequence of a partial murine SPG4 polypeptide.
XX
KW   Mouse; SPG4 gene; spastin; PSF-AD; gene therapy;
KW   autosomal dominant familial spastic paraplegia; ss.
XX
OS   Mus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..1517
FT                   /*tag=  a
FT                   /product= "spastin"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
DR   P-PSDB; AAB68136.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening  -
XX
PS   Claim 2; Page 123; 145pp; French.
XX
CC   The present sequence encodes a partial spastin polypeptide. Spastin is
CC   encoded by the SPG4 gene. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 1689 BP; 525 A; 357 C; 416 G; 391 T; 0 other;

  Query Match             14.6%;  Score 73;  DB 22;  Length 1689;
  Best Local Similarity   93.8%;  Pred. No. 7.3e-06;
  Matches   76;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        392 AGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGC 451
              ||||:||| ||:||| |:||| :||| :|.||| | |||: |||:|||||:|| ||:
Db          1 AGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCTTCGAGTACATCTCCATTGCCCTGC 60

Qy        452 GCATCGATGAGGATGAGAAAG 472
              ||||:|| ||||: |||||'|
Db         61 GCATCGACGAGGAAGAGAAAG 81


RESULT 12
ABS71875
ID   ABS71875 standard; DNA; 500 BP.
XX
AC   ABS71875;
XX
DT   02-DEC-2002  (first entry)
XX
DE   Human GTP-Rho binding protein 2 single-exon probe #16.
XX
KW   Human; ss; GTP-Rho binding protein 2; GRBP2; chromosome 19q12;
KW   oncogene; tumour; liposarcoma; ichthyosis congenita III; probe;
KW   benign familial infantile convulsion; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1231216-A2.
XX
PD   14-AUG-2002.
XX
PF   17-JAN-2002; 2002EP-0001026.
XX
PR   30-JAN-2001; 2001WO-US00663.
PR   30-JAN-2001; 2001WO-US00664.
PR   30-JAN-2001; 2001WO-US00665.
PR   30-JAN-2001; 2001WO-US00666.
PR   30-JAN-2001; 2001WO-US00667.
PR   30-JAN-2001; 2001WO-US00668.
PR   30-JAN-2001; 2001WO-US00669.
PR   30-JAN-2001; 2001WO-US00670.
PR   29-JUN-2001; 2001US-0895040.
XX
PA   (AEOM-) AEOMICA INC.
XX
PI   Shannon ME,  Ol Y;
XX
DR   WPI; 2002-684026/74.
XX
PT   Novel GTP-Rho binding protein 2 and nucleic acids encoding the protein,
PT   useful for the manufacture of a medicament for treating a disease
PT   associated with altered expression or activity of human GRBP2 protein
PT   -
XX
PS   Example 4; Page 57-58; 101pp; English.
XX
CC   The invention relates to an isolated GTP-Rho binding protein 2 (GRBP2)
CC   polypeptide or a fragment of at least 6 amino acids or a sequence in
CC   which at least 95% of deviations from GRBP2 sequences are conservative
CC   substitutions. Also included are an isolated nucleic acid (GRBP2 NA)
CC   encoding GRBP2 comprising the full length cDNA or CDS, fragments or
CC   variants, GRBP2 vectors, host cells, antibodies, transgenic non-human
CC   animals modified to contain GRBP2 NA (or unable to express the endogenous
CC   orthologue of GRBP2), diagnosing a disease caused by a mutation in human
CC   GRBP2 or altered expression of GRBP2, ant-agonists of GRBP2, GRBP2
CC   microarrays, fusion proteins and screening for agents that modulate the
CC   expression of GRBP2 NA. GRBP2 is useful for identifying binding partners
CC   of GRBP2. GRBP2, GRBP2 NA and Ab are useful in therapy and in the
CC   manufacture of a medicament for the treatment or prevention of a disorder
CC   associated with increased or decreased expression or activity of human
CC   GRBP2 (e.g. tumours, liposarcoma, ichthyosis congenita III and benign
CC   familial infantile convulsion, all associated with the chromosomal
CC   location of GRBP2, 19q12). GRBP2 is useful as a standard in immunoassay
CC   specific for the proteins, to be used in a therapeutic agent, as
CC   vaccines, to be and as antigens (e.g. for epitope mapping) or immunogens
CC   (e.g. for raising antibodies. GRBP2 NA is useful as hybridisation probes,
CC   to prime synthesis of nucleic acids, to prime first strand cDNA sequence
CC   on an mRNA template, and to drive in vivo expression of the proteins. The
CC   vector is useful for shuttling GRBP2 NA between host cells derived from
CC   disparate organisms, for inserting GRBP2 NA into host cell chromosome,
CC   for expressing sense or antisense RNA transcripts of GRBP2 NA in vitro or
CC   within a host cell, and for expressing GRBP2 alone or as fusions to
CC   heterologous polypeptides. The antibody is useful as an analytical
CC   reagent for detection and quantification of GRBP2 and as an immuno
CC   therapeutic agent and is useful for flow cytometric detection, for
CC   scanning laser cytometric detection, or for fluorescent immunoassay.
CC   The present sequence is a single exon probe for GRBP2.
XX
SQ   Sequence 500 BP; 55 A; 163 C; 228 G; 54 T; 0 other;

  Query Match             10.1%;  Score 50.8;  DB 24;  Length 500;
  Best Local Similarity   45.5%;  Pred. No. 0.2;
  Matches  181;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;
```

```
Qy          2 GTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGA 61

Db         68 GTGCTCGGCCTCAGGCGGGGCTAGAAGGGCCGCGGGACGGGGTGGGAGTGGAGGGGCGGG 127

Qy         62 ATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTC 121

Db        128 GAAGGGCGGGGACAGGGGCGGGGCCGCACGTCCTCTCGGGCCAGCCTCAGCCGCCGCGCC 187

Qy        122 CCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGC 181

Db        188 TCAGTCCGCCGTCCGCCCTCCGCGCCCGCGCCGCTAGCATGACCGACGGCGTGTTGCCCG 247

Qy        182 CCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCG 241

Db        248 CGGCCCCCCAGCCGCTGGAGAAGGAGAACGACGGCTACTTTCGGAAGGTGGGGCGCTTCG 307

Qy        242 CGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCT 301

Db        308 GGCGGGCGGGCGGGGACCTGCGGGCTCCAGACCTCCTTTCCCCCGGGCCCTGCAGGGGCC 367

Qy        302 TCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGG 361

Db        368 GTGGGAGCGCGCAGTCGGGGTCCACTAGGCCGGGAGGGGGAGGGGGCGCACTGGGCCGGC 427

Qy        362 CCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAG 399

Db        428 GCTGGCCGGACGGAGGCTGGCGGGGAGGAGTGGGGGCG 465


RESULT 13
AAI71296
ID   AAI71296 standard; DNA; 2795 BP.
XX
AC   AAI71296;
XX
DT   24-JAN-2002  (first entry)
XX
DE   Streptomyces griseus RppA protein encoding DNA SEQ ID NO:1.
XX
KW   Streptomyces griseus; RppA; pyran-2-one; drug; agricultural; ds.
XX
OS   Streptomyces griseus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1596..2714
FT                   /*tag=  a
FT                   /product= "RppA protein"
XX
PN   JP2001238677-A.
XX
PD   04-SEP-2001.
XX
PF   01-MAR-2000; 2000JP-0055782.
XX
PR   01-MAR-2000; 2000JP-0055782.
XX
PA   (MITR ) MITSUBISHI RAYON CO LTD.
XX
DR   WPI; 2001-613891/71.
DR   P-PSDB; AAG66148.
XX
PT   Preparation of pyran-2-one derivative, used as intermediate and as
PT   additive for thermosensible transfer paper, comprises contacting
PT   carbonyl compound with microbe containing polyketide synthase gene -
XX
PS   Claim 6; Page 7-8; 11pp; Japanese.
XX
CC   The present invention describes the preparation of pyran-2-one
CC   derivatives. The preparation method comprises contacting carbonyl
CC   compounds with microbes containing a polyketide synthase gene.
CC   Pyran-2-one derivatives can be used as synthetic intermediates for
CC   drugs and agricultural chemicals. They can also be used as additives
CC   for thermosensible transfer paper. The present sequence encodes the
CC   specifically claimed Streptomyces griseus RppA protein, which is used
CC   in the exemplification of the present invention.
XX
SQ   Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;

  Query Match             9.9%;  Score 49.6;  DB 22;  Length 2795;
  Best Local Similarity   50.9%;  Pred. No. 0.37;
  Matches  118;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        235 GGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGC 294

Db        779 GTCTCCGGGCCGCTGCGGCTGATCACCATCTGCGAGGCACTCGACATCCCCGAGGCGGAC 838

Qy        295 CAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354

Db        839 CGGCCCTGGCTCCGGGCCCACGCCATGACCATGATGAACGTCGGGGGCCGCGGGCAAGCAG 898

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTC 414

Db        899 GACGCGGTGCCGCGCCAAGGCGGAGCTGCGCGGCTACTTCCAGGAGCTGACCGCCGACCGC 958

Qy        415 CACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATG 466

Db        959 CGCCGCTCCCCGGGCGAGGACCTCATCAGCACCCTGGCCACCGCCCGGGACG 1010


RESULT 14
AAF81954
ID   AAF81954 standard; DNA; 2795 BP.
XX
AC   AAF81954;
XX
DT   18-JUN-2001  (first entry)
XX
DE   Streptomyces griseus actinomycetes polyketide synthase encoding DNA.
XX
KW   Streptomyces griseus; actinomycete; polyketide synthase; rppA; drug;
KW   1,3,6,8-tetrahydroxynaphthalene; melanine; agricultural chemical;
KW   malonyl CoA; ds.
XX
OS   Streptomyces griseus.
XX
FH   Key             Location/Qualifiers
FT   CDS             1596..2714
FT                   /*tag=  a
FT                   /product= "Actinomycetes polyketide synthase"
XX
PN   JP2000342269-A.
XX
PD   12-DEC-2000.
XX
PF   02-JUN-1999;   99JP-0155526.
XX
PR   02-JUN-1999;   99JP-0155526.
XX
PA   (MITR ) MITSUBISHI RAYON CO LTD.
XX
DR   WPI; 2001-303158/32.
DR   P-PSDB; AAB74843.
XX
PT   Preparation of 1,3,6,8-tetrahydroxynaphthalene for use as a synthetic
PT   material for melanine and as an intermediate for drugs and agricultural
PT   chemicals, comprises using a microbe and an enzyme -
XX
PS   Claim 2; Page 6-7; 10pp; Japanese.
XX
CC   The present invention describes a method for preparing
CC   1,3,6,8-tetrahydroxynaphthalene comprising using a microbe cell
CC   containing an rppA gene encoding actinomycetes polyketide synthase, to
CC   convert malonyl CoA as the substrate or a substance which can produce
CC   malonyl CoA in the microbe cell to 1,3,6,8-tetrahydroxynaphthalene and
CC   recovering it. 1,3,6,8-tetrahydroxynaphthalene is used as a synthetic
CC   material for melanine and as an intermediate for drugs and agricultural
```

```
CC   chemicals. The present sequence encodes the actinomycetes polyketide
CC   synthase protein isolated from Streptomyces griseus.
XX
SQ   Sequence 2795 BP; 435 A; 1051 C; 908 G; 401 T; 0 other;

  Query Match             9.9%;  Score 49.6;  DB 22;  Length 2795;
  Best Local Similarity   50.9%;  Pred. No. 0.37;
  Matches  118;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy        235 GGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGC 294

Db        779 GTCTCCGCGCCGCTGCCGCTGATCACCATCTGCGAGGCACTCGACATCCCCGAGGCGGAC 838

Qy        295 CAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354

Db        839 CGGCCCTGGCTCCGGGCCCACGCCATGACCATGATGAACGTCGGGGCCGCGGGCAAGCAG 898

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTC 414

Db        899 GACGCGGTGCGCGCCAAGGCGGAGCTGCGCGGCTACTTCCAGGAGCTGACCGCCGACCGC 958

Qy        415 CACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATG 466

Db        959 CGCCGCTCCCCGGGCGAGGACCTCATCAGCACCCTGGCCACCGCCCGGGACG 1010

RESULT 15
AAT33617
ID   AAT33617 standard; DNA; 1269 BP.
XX
AC   AAT33617;
XX
DT   03-JAN-1997  (first entry)
XX
DE   DxrA gene encoding daunorubicin 14-hydroxylase.
XX
KW   dxrA; doxorubicin; bioconversion; anthracycline; antitumour therapy;
KW   daunorubicin 14-hydroxylase; ds.
XX
OS   Streptomyces peucetius strain 29050.
XX
PN   WO9627014-A1.
XX
PD   06-SEP-1996.
XX
PF   20-FEB-1996;   96WO-EP00692.
XX
PR   27-FEB-1995;   95US-0396216.
XX
PA   (PHAA ) PHARMACIA SPA.
XX
PI   Breme U,  Colombo AL,  Hutchinson CR,  Otten S,  Scotti C;
PI   Solari INVENTI A, Inventi AS;
XX
DR   WPI; 1996-412780/41.
DR   P-PSDB; AAW00729.
XX
PT   New DNA encoding daunorubicin 14-hydroxylase - useful for converting
PT   daunorubicin to the antitumour agent doxorubicin
XX
PS   Claim 6; Page 15-18; 31pp; English.
XX
CC   The present sequence is the drxA gene coding sequence for daunorubicin
CC   14-hydroxylase, isolated from Streptomyces peucetius strain 29050. The
CC   enzyme converts daunorubicin to doxorubicin, an anthracycline used widely
CC   as an antitumour agent. A vector contg. the drxA gene could be used to
CC   transform a suitable host cell which produces daunorubicin, in order to
CC   enhance bioconversion of daunorubicin to doxorubicin. Bioconversion can
CC   be carried out either by using directly the free or immobilised
CC   transformed cells or by isolating the enzyme, which can be used in the
CC   free form or immobilised to resins, glass, cellulose or similar
CC   substances by ionic or covalent bonds, or grafted to fibres permeable to
CC   the substrate or insolubilised by cross-linkage.
XX
SQ   Sequence 1269 BP; 164 A; 484 C; 442 G; 179 T; 0 other;

  Query Match             9.9%;  Score 49.4;  DB 17;  Length 1269;
  Best Local Similarity   48.4%;  Pred. No. 0.4;
  Matches  137;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167

Db        360 CGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGC 419

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227

Db        420 CTCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 479

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287

Db        480 GTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGC 539

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347

Db        540 CGTCAGCGTTCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGGTGACGGCAC 599

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390

Db        600 GGACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGC 642

Search completed: October 24, 2003, 18:47:17
Job time : 128.251 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 13:34:49 ; Search time 31.3542 Seconds
(without alignments)
7052.739 Million cell updates/sec

Title: US-09-830-902-1_COPY_10000_10500
Perfect score: 501
Sequence: 1 ggttcccgtcggtctgcggg.....gggctgggggagggggcggc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	56.6	11.3	7218	1	US-08-232-463-14	Sequence 14, Appl
2	49.4	9.9	1269	1	US-08-396-218-1	Sequence 1, Appli
3	49.4	9.9	1269	1	US-08-760-116-1	Sequence 1, Appli
4	47.8	9.5	1735	1	US-08-122-863-10	Sequence 10, Appl
5	47.8	9.5	1735	5	PCT-US92-10885-10	Sequence 10, Appl
6	46.4	9.3	12001	1	US-08-458-568A-11	Sequence 11, Appl
7	46.2	9.2	2081	2	US-09-096-982-7	Sequence 7, Appli
8	46.2	9.2	2081	2	US-08-653-650A-7	Sequence 7, Appli
9	46.2	9.2	3013	2	US-09-096-982-6	Sequence 6, Appli
10	46.2	9.2	3013	2	US-08-653-650A-6	Sequence 6, Appli
11	46.2	9.2	3196	2	US-09-096-982-4	Sequence 4, Appli
12	46.2	9.2	3196	2	US-08-653-650A-4	Sequence 4, Appli
13	46	9.2	803	1	US-07-928-611-12	Sequence 12, Appl
14	46	9.2	803	2	US-08-487-811A-12	Sequence 12, Appl
15	46	9.2	803	3	US-09-060-694-12	Sequence 12, Appl
16	46	9.2	803	4	US-09-378-074-12	Sequence 12, Appl
17	46	9.2	803	5	PCT-US93-07370-12	Sequence 12, Appl
18	46	9.2	1504	4	US-09-016-434-1276	Sequence 1276, Ap
19	46	9.2	1610	1	US-08-056-051-5	Sequence 5, Appli
20	46	9.2	1610	1	US-07-928-611-21	Sequence 21, Appl
21	46	9.2	1610	2	US-08-487-811A-21	Sequence 21, Appl
22	46	9.2	1610	3	US-09-060-694-21	Sequence 21, Appl
23	46	9.2	1610	4	US-09-378-074-21	Sequence 21, Appl
24	46	9.2	1610	5	PCT-US93-07370-21	Sequence 21, Appl
25	45.2	9.0	1877	4	US-09-780-173A-10	Sequence 10, Appl
26	45	9.0	3177	1	US-08-042-747A-4	Sequence 4, Appli
c 27	44	8.8	8438	1	US-07-945-283-1	Sequence 1, Appli
c 28	44	8.8	4403765	3	US-09-103-840A-2	Sequence 2, Appli
c 29	44	8.8	4411529	3	US-09-103-840A-1	Sequence 1, Appli
c 30	43.2	8.6	1584	4	US-09-252-991A-14590	Sequence 14590, A
31	43.2	8.6	1995	4	US-09-252-991A-15174	Sequence 15174, A
32	43.2	8.6	2109	4	US-09-252-991A-15038	Sequence 15038, A
33	43	8.6	2064	1	US-08-343-428-1	Sequence 1, Appli
c 34	42.6	8.5	2150	2	US-08-318-837-1	Sequence 1, Appli
35	42.2	8.4	4257	2	US-08-690-473-1	Sequence 1, Appli
36	42.2	8.4	4257	3	US-09-259-821A-1	Sequence 1, Appli
37	42.2	8.4	4257	3	US-08-843-659-1	Sequence 1, Appli
c 38	42.2	8.4	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 39	42	8.4	1413	4	US-08-984-709A-52	Sequence 52, Appl
40	42	8.4	11219	1	US-07-642-734C-1	Sequence 1, Appli
41	42	8.4	11219	3	US-08-439-009A-1	Sequence 1, Appli
42	41.8	8.3	34094	4	US-09-292-034-1	Sequence 1, Appli
43	41.8	8.3	152331	3	US-09-128-155-16	Sequence 16, Appl
c 44	41.6	8.3	441	4	US-09-252-991A-3998	Sequence 3998, Ap
45	41.6	8.3	2061	4	US-09-252-991A-3938	Sequence 3938, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: CORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 11.3%; Score 56.6; DB 1; Length 7218;

```
  Best Local Similarity   7.5%;  Pred. No. 0.0011;
  Matches   32;  Conservative  219;  Mismatches  178;  Indels    0;  Gaps    0;

Qy         28 TTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGAATTCTCCGGGTGGACGAGGGAAGAAG 87

Db        996 TTTTTCCTTCGTTTGCCATACGCTCACAGAATTAATTCCGAGCTTGGCTGCAGGTCGAGG 1055

Qy         88 AAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCC 147

Db       1056 GAGCTTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1115

Qy        148 CCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTG 207

Db       1116 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1175

Qy        208 TACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCAC 267

Db       1176 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1235

Qy        268 CTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAG 327

Db       1236 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1295

Qy        328 AGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCGGCGCCGGTGCCGGGC 387

Db       1296 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1355

Qy        388 GGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCC 447

Db       1356 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1415

Qy        448 CTGCGCATC 456

Db       1416 YYYYYYYYY 1424
```

RESULT 2
US-08-396-218-1
; Sequence 1, Application US/08396218
; Patent No. 5695966
; GENERAL INFORMATION:
;   APPLICANT: INVENTI, Augusto
;   APPLICANT: BREME, Umberto
;   APPLICANT: COLOMBO, Anna L
;   APPLICANT: HUTCHINSON, Charles R
;   APPLICANT: OTTEN, Sharee
;   APPLICANT: SCOTTI, Claudio
;   TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM
;   STREET: 655 Fifteenth Street, N. W., Suite 330 - G
;   STREET: Street Lobby
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005-5701
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/396,218
;   FILING DATE:
;   CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;   NAME: KITTS, Monica C
;   REGISTRATION NUMBER: 36,105
;   REFERENCE/DOCKET NUMBER: P1615-5002
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202/638-5000
;   TELEFAX: 202/638-4810

; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1269 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Streptomyces peucetius
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1269
US-08-396-218-1

```
  Query Match             9.9%;  Score 49.4;  DB 1;  Length 1269;
  Best Local Similarity   48.4%;  Pred. No. 0.03;
  Matches  137;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCGCCCCTCCCGCCGCCGG 167

Db        360 CGAGCGGACGGATCGCATCGCCGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGC 419

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227

Db        420 CTCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 479

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287

Db        480 GTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGC 539

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347

Db        540 CGTCAGCGTTCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGGCGGGGTGACGGCAC 599

Qy        348 AGCACCTGCCTCGGCCTCGGCCCGGCGCCGGTGCCGGGCGGC 390

Db        600 GGACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGC 642
```

RESULT 3
US-08-760-116-1
; Sequence 1, Application US/08760116
; Patent No. 5786190
; GENERAL INFORMATION:
;   APPLICANT: INVENTI, Augusto
;   APPLICANT: BREME, Umberto
;   APPLICANT: COLOMBO, Anna L
;   APPLICANT: HUTCHINSON, Charles R
;   APPLICANT: OTTEN, Sharee
;   APPLICANT: SCOTTI, Claudio
;   TITLE OF INVENTION: PROCESS FOR PREPARING DOXORUBICIN
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: NIKAIDO, MARMELSTEIN, MURRAY & ORAM LLP
;   STREET: 655 Fifteenth Street, N. W., Suite 330 - G
;   STREET: Street Lobby
;   CITY: Washington
;   STATE: DC
;   COUNTRY: USA
;   ZIP: 20005-5701
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/760,116
;   FILING DATE: 3-DEC-1996
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/396,218

```
;      FILING DATE:  27-FEB-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KITTS, Monica C
;      REGISTRATION NUMBER:  36,105
;      REFERENCE/DOCKET NUMBER:  P1615-6007
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202/638-5000
;      TELEFAX:  202/638-4810
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1269 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptomyces peucetius
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1269
US-08-760-116-1

  Query Match             9.9%;  Score 49.4;  DB 1;  Length 1269;
  Best Local Similarity   48.4%;  Pred. No. 0.03;
  Matches  137;  Conservative    0;  Mismatches  146;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCGCCGCGCGG 167
Db        360 CGAGCGGACGGATCGCATCGCCGGCGATCGCCGGCCGGCTGCTCACCGAACTCGCCGACGC 419

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db        420 CTCCGGCCGGTCGGGCAAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 479

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGG 287
Db        480 GTTGGTCATCTGCGAGCTGCTCGGTGTGCCGGTCACCGATCCGGCGATGGCCCGCGAGGC 539

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCAAGAGGAGCTCCGGGGCCGCGCCAG 347
Db        540 CGTCAGCGTTCTCAAGGCACTCGGCCTCGGCGGCCGCAGAGCGGCGGGGGTGACGGCAC 599

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGCGGC 388
Db        600 GGACCCTGCCGGGGGCGTGCCGGACACCTCGGCCCTGGAGAGC 642


RESULT 4
US-08-102-863-10
; Sequence 10, Application US/08102863
; Patent No. 5466590
;  GENERAL INFORMATION:
;    APPLICANT:  SARIASLANI, SIMA
;    TITLE OF INVENTION:  CONSTITUTIVE
;    TITLE OF INVENTION:  EXPRESSION OF P450SOY
;    TITLE OF INVENTION:  AND FERREDOXIN-SOY IN
;    TITLE OF INVENTION:  STREPTOMYCES
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  USA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,
;      SOFTWARE:  Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/102,863
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/807,001
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  GALLEGOS, R. THOMAS
;      REGISTRATION NUMBER:  32,692
;      REFERENCE/DOCKET NUMBER:  CR-9000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  302-892-7342
;      TELEFAX:  302-892-7949
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1735 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-102-863-10

  Query Match             9.5%;  Score 47.8;  DB 1;  Length 1735;
  Best Local Similarity   47.0%;  Pred. No. 0.068;
  Matches  148;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        127 CCTCCGCCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCGCGGGCCGGCCCCTCCGCCCGAG 186
Db        205 CCGGCCCGCCAGAACCTCGACCCCACCTCCCCGGCCCCCGCGACGTCCTTCCCGCAGGAC 264

Qy        187 TCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTG 246
Db        265 CGCGGGTGCCCCTACCACCCGCCCGCCGGGTACGCACCGCTGCGCGAGGGCCGCCCGCTG 324

Qy        247 CTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCC 306
Db        325 AGCCGGGTCACCCTCTTCGACGGACGCCCGGTCTGGGCGGTCACCGGCACGCCGTGGCC 384

Qy        307 CGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCG 366
Db        385 CGTCGGCTACTGGCGGACCCGCGGCTCTCCACCGACCGCAGCCACCCGGACTTCCCCGTC 444

Qy        367 GCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCC 426
Db        445 CCGGCCGAGCGGTTCGCCGGCGCGCAGCGGCGCCGCGTCGCTCTGCTCGGCGTCGACGAC 504

Qy        427 TTCGAGTACATCTCC 441
Db        505 CCCGAGCACAACACC 519


RESULT 5
PCT-US92-10885-10
; Sequence 10, Application PC/TUS9210885
;  GENERAL INFORMATION:
;    APPLICANT:  SARIASLANI, SIMA
;    TITLE OF INVENTION:  CONSTITUTIVE
;    TITLE OF INVENTION:  EXPRESSION OF P450SOY
;    TITLE OF INVENTION:  AND FERREDOXIN-SOY IN
;    TITLE OF INVENTION:  STREPTOMYCES
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  E. I. DU PONT DE NEMOURS
;      ADDRESSEE:  AND COMPANY
;      STREET:  1007 MARKET STREET
;      CITY:  WILMINGTON
;      STATE:  DELAWARE
;      COUNTRY:  USA
;      ZIP:  19898
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch,
```

```
;       MEDIUM TYPE:  1.0 MB
;       COMPUTER:  Macintosh
;       OPERATING SYSTEM:  Macintosh System, 6.0
;       SOFTWARE:  Microsoft Word, 4.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US92/10885
;       FILING DATE:  19921216
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  GALLEGOS, R. THOMAS
;       REGISTRATION NUMBER:  32,692
;       REFERENCE/DOCKET NUMBER:  CR-9000-A
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  302-892-7342
;       TELEFAX:  302-892-7949
;   INFORMATION FOR SEQ ID NO:  10:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1735 base pairs
;       TYPE:  NUCLEIC ACID
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
PCT-US92-10885-10

  Query Match             9.5%;  Score 47.8;  DB 5;  Length 1735;
  Best Local Similarity   47.0%;  Pred. No. 0.068;
  Matches  148;  Conservative    0;  Mismatches  167;  Indels    0;  Gaps    0;

Qy        127 CCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAG 186
Db        205 CCGGCCCGCCAGAACCTCGACCCCACCTCCCCGGCCCCCGCGACGTCCTTCCCGCAGGAC 264

Qy        187 TCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTG 246
Db        265 CGCGGGTGCCCCTACCACCCGCCCGCCGGGTACGCACCGCTGCGCGAGGGCCGCCCGCTG 324

Qy        247 CTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCC 306
Db        325 AGCCGGGTCACCCTCTTCGACGGACGCCCGGTCTGGGCGGTCACCGGGCACGCCCTGGCC 384

Qy        307 CGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCG 366
Db        385 CGTCGGCTACTGGCGGACCCGCGGCTCTCCACCGACCGCAGCCACCCGGACTTCCCCGTC 444

Qy        367 GCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCC 426
Db        445 CCGGCCGAGCGGTTCGCCGGCGCGCAGCGGCGCCGCGTCGCTCTGCTCGGCGTCGACGAC 504

Qy        427 TTCGAGTACATCTCC 441
Db        505 CCCGAGCACAACACC 519

RESULT 6
US-08-458-568A-11
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11

  Query Match             9.3%;  Score 46.4;  DB 1;  Length 12001;
  Best Local Similarity   51.3%;  Pred. No. 0.16;
  Matches  158;  Conservative    0;  Mismatches  146;  Indels    4;  Gaps    2;

Qy        111 CCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCC 170
Db       1258 CCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCC 1317

Qy        171 GGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTT 230
Db       1318 CCCCGCTCCTCCCCCCGCTCCTCCC--CCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCT 1375

Qy        231 TGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCT 290
Db       1376 CCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCTCCCCCCGCTCCT 1435

Qy        291 CTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGC 350
Db       1436 C--CCCCCGCTCCTCCCCCCGCTCCCGCGGCCCCGCCCCCCACGCCCGCCGCGCGCGCGC 1493

Qy        351 ACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGT 410
Db       1494 ACGCCGCCCGGACCGCCGCCCCGCCTTTTTTGCGCGCGCGCGCGCCCGCGGGGGGCCCGGG 1553

Qy        411 CTTCCACA 418
Db       1554 CTGCCACA 1561

RESULT 7
US-09-096-982-7
; Sequence 7, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
```

```
;       CITY:  Cleveland
;       STATE:  Ohio
;       COUNTRY:  USA
;       ZIP:  44114-2688
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/096,982
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Golrick, Mary E.
;       REGISTRATION NUMBER:  34829
;       REFERENCE/DOCKET NUMBER:  22727/00131
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  216-622-8458
;       TELEFAX:  216-241-0816
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2081 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  227..1649
US-09-096-982-7

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 2081;
  Best Local Similarity   47.7%;  Pred. No. 0.15;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        742 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 801

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        802 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 861

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        862 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 921

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        922 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 981

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        982 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 1024


RESULT 8
US-08-653-650A-7
; Sequence 7, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,650A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2081 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  227..1649
US-08-653-650A-7

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 2081;
  Best Local Similarity   47.7%;  Pred. No. 0.15;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        742 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 801

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        802 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 861

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        862 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 921

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        922 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 981

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db        982 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 1024


RESULT 9
US-09-096-982-6
; Sequence 6, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
```

```
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/096,982
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Golrick, Mary E.
;       REGISTRATION NUMBER:  34829
;       REFERENCE/DOCKET NUMBER:  22727/00131
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  216-622-8458
;       TELEFAX:  216-241-0816
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3013 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  other nucleic acid
US-09-096-982-6

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 3013;
  Best Local Similarity   47.7%;  Pred. No. 0.16;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |||||  |
Db       1674 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 1733

Qy        168 GCCGGCCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       1734 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 1793

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              |||  |  ||  |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       1794 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 1853

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              |  ||  |  |  |  ||  |  |  |  ||||  |  ||  |  |
Db       1854 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 1913

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
              |  ||||  ||  |  |  |  ||  |  |  |  |  |  |
Db       1914 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 1956


RESULT 10
US-08-653-650A-6
; Sequence 6, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,650A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3013 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid
US-08-653-650A-6

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 3013;
  Best Local Similarity   47.7%;  Pred. No. 0.16;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |||||  |
Db       1674 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 1733

Qy        168 GCCGGCCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       1734 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 1793

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              |  |  ||  |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |
Db       1794 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 1853

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              ||  ||  |  |  ||  |  |  |  ||||  |  ||  |  |
Db       1854 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 1913

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
              |  ||||  ||  |  |  |  ||  |  |  |  |  |  |
Db       1914 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 1956


RESULT 11
US-09-096-982-4
; Sequence 4, Application US/09096982
; Patent No. 5962293
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/096,982
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3196 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1498..2764
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1498..2764
US-09-096-982-4

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 3196;
  Best Local Similarity   47.7%;  Pred. No. 0.16;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db       1857 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 1916

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db       1917 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 1976

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db       1977 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 2036

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db       2037 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 2096

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
Db       2097 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 2139

RESULT 12
US-08-653-650A-4
; Sequence 4, Application US/08653650A
; Patent No. 5976830
;  GENERAL INFORMATION:
;    APPLICANT:  Strohl, William R.
;    APPLICANT:  Dickens, Michael L.
;    APPLICANT:  DeSanti, Charles L.
;    TITLE OF INVENTION:  METHOD OF PRODUCING DOXORUBICIN
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CALFEE, HALTER & GRISWOLD
;      STREET:  800 Superior Avenue, Suite 1400
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  USA
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/653,650A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22727/00131
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  216-622-8458
;      TELEFAX:  216-241-0816
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3196 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  mat_peptide
;      LOCATION:  1498..2764
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1498..2764
US-08-653-650A-4

  Query Match             9.2%;  Score 46.2;  DB 2;  Length 3196;
  Best Local Similarity   47.7%;  Pred. No. 0.16;
  Matches  135;  Conservative    0;  Mismatches  148;  Indels    0;  Gaps    0;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db       1857 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGCTCACCGAACTCGCCGACTC 1916

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db       1917 CTCCGACCGGTCGGGCGAACCGGCCGAGCTGATCGGCGGCTTCGCGTACCACTTCCCGCT 1976

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db       1977 GTTGGTCATCTGCGAACTGCTCGGCGTGCCGGTCACCGATCCGGCAATGGCCCGCGAGGC 2036

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db       2037 CGTCGGCGTGCTCAAGGCACTCGGCCTCGGCGGCCCGCAGAGCGCCGGCGGTGACGGCAC 2096

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
Db       2097 GGACCCTGCCGGGGACGTGCCGGACACGTCGGCGCTGGAGAGC 2139

RESULT 13
US-07-928-611-12
; Sequence 12, Application US/07928611
; Patent No. 5569601
;  GENERAL INFORMATION:
;    APPLICANT:  Van Tol, Hubert H.M.
;    APPLICANT:  Civelli, Olivier
;    TITLE OF INVENTION:  A No. 556960lel Human Dopamine Receptor and Uses
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Allegretti & Witcoff, Ltd.
;      STREET:  10 South Wacker Drive, Suite 3000
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/928,611
;      FILING DATE:  19920810
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 556960lnan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  9C,1092-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-715-1000
```

```
;      TELEFAX:  312-715-1234
;      TELEX:  810-221-8317
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1..803
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "Alternate Exon 3: D4.7"
;      OTHER INFORMATION:  /note= "This sequence represents the third exon of
;      OTHER INFORMATION:  allele D4.7 of the human D4 dopamine receptor
;      OTHER INFORMATION:  gene"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  257..262
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "PstI site"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "PstI site"
;      OTHER INFORMATION:  /label= PstI
;      OTHER INFORMATION:  /note= "This sequence is a PstI site whereby
;      OTHER INFORMATION:  digestion of human genomic DNA produces a RFLP"
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  346..682
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /rpt_type= "tandem"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /rpt_unit= 346 .. 394
;      OTHER INFORMATION:  /note= "This sequence is a repeat found in 7 known
;      OTHER INFORMATION:  alleles of the human D4 dopamine receptor gene
;      OTHER INFORMATION:  encoding a 16 amino acid sequence repeated 7 times
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..803
US-07-928-611-12

  Query Match             9.2%;  Score 46;  DB 1;  Length 803;
  Best Local Similarity   49.7%;  Pred. No. 0.15;
  Matches  145;  Conservative    0;  Mismatches  145;  Indels    2;  Gaps    1;

Qy        117 GCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCC 176
Db        297 GCCGCGCGCCCCGCCGACCCAGCGGCCCTGGCCCGCCTTCCCCCACGCCACCCGCGCCCC 356

Qy        177 TCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGG 236
Db        357 GCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCCGCGCCCGGCCTTCCCCGGG 416

Qy        237 CTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296
Db        417 GTCCCTGCGGCCCCGACTGTGCGCCCGCCGCGCCCGGCCTCCCCCGGACCCCTGCGGCC 476

Qy        297 GCGCTTCTCCCGCGCCCTCAT--GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354
Db        477 CCGACTGTGCGCCCCCCGCGCCCGGCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGC 536

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCC 406
Db        537 CCCCCGCGCCCGGCCTTCCCCGGGGTCCCTGCGGCCCCGACTGTGCGCCCCC 588


RESULT 14
US-08-487-811A-12
; Sequence 12, Application US/08487811A
; Patent No. 5883226
;  GENERAL INFORMATION:
;    APPLICANT:  Civelli, Olivier
;    APPLICANT:  Van Tol, Hubert H.M.
;    TITLE OF INVENTION:  A No. 5883226el Human Dopamine Receptor and Uses
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  McDonnell Boehnen Hulbert & Berghoff
;      STREET:  300 South Wacker Drive
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,811A
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5883226nan, Kevin E
;      REGISTRATION NUMBER:  35,303
;      REFERENCE/DOCKET NUMBER:  90,1092-D
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-913-0001
;      TELEFAX:  312-913-0002
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  803 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  exon
;      LOCATION:  1..803
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "Alternate Exon 3: D4.7"
;      OTHER INFORMATION:  /note= "This sequence represents the third exon of
;      OTHER INFORMATION:  allele D4.7 of the human D4 dopamine receptor
;      OTHER INFORMATION:  gene"
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  257..262
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /function= "PstI site"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /standard_name= "PstI site"
;      OTHER INFORMATION:  /label= PstI
;      OTHER INFORMATION:  /note= "This sequence is a PstI site whereby
;      OTHER INFORMATION:  digestion of human genomic DNA produces a RFLP"
;    FEATURE:
;      NAME/KEY:  repeat_region
;      LOCATION:  346..682
;      IDENTIFICATION METHOD:  experimental
;      OTHER INFORMATION:  /rpt_type= "tandem"
;      OTHER INFORMATION:  /evidence= EXPERIMENTAL
;      OTHER INFORMATION:  /rpt_unit= 346 .. 394
;      OTHER INFORMATION:  /note= "This sequence is a repeat found in 7 known
;      OTHER INFORMATION:  alleles of the human D4 dopamine receptor gene
;      OTHER INFORMATION:  encoding a 16 amino acid sequence repeated 7 times
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2..803
US-08-487-811A-12

  Query Match             9.2%;  Score 46;  DB 2;  Length 803;
  Best Local Similarity   49.7%;  Pred. No. 0.15;
  Matches  145;  Conservative    0;  Mismatches  145;  Indels    2;  Gaps    1;
```

```
Qy        117 GCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCC 176
Db        297 GCCGCGCGCCCCGCCGACCCAGCGGCCCTGGCCCGCCTTCCCCCACGCCACCCGCGCCCC 356

Qy        177 TCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGG 236
Db        357 GCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCCGCGCCCGGCCTTCCCCGGG 416

Qy        237 CTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296
Db        417 GTCCCTGCGGCCCCGACTGTGCGCCCGCCGCGCCCGGCCTCCCCCGGACCCCTGCGGCC 476

Qy        297 GCGCTTCTCCCGCGCCCTCAT--GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354
Db        477 CCGACTGTGCGCCCCCCGCGCCCGGCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGC 536

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCC 406
Db        537 CCCCCGCGCCCGGCCTTCCCCGGGGTCCCTGCGGCCCCGACTGTGCGCCCCC 588
```

RESULT 15
US-09-060-694-12
; Sequence 12, Application US/09060694
; Patent No. 6203998
; GENERAL INFORMATION:
;   APPLICANT: Civelli, Olivier
;   APPLICANT: Van Tol, Hubert H.M.
;   TITLE OF INVENTION: A No. 6203998el Human Dopamine Receptor and Uses
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
;   STREET: 300 South Wacker Drive
;   CITY: Chicago
;   STATE: IL
;   COUNTRY: USA
;   ZIP: 60606
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/060,694
;   FILING DATE: 15-APR-1998
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 6203998nan, Kevin E
;   REGISTRATION NUMBER: 35,303
;   REFERENCE/DOCKET NUMBER: 90,1092-MM
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-913-0001
;   TELEFAX: 312-913-0002
;   TELEX:
; INFORMATION FOR SEQ ID NO: 12:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 803 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1..803
;   IDENTIFICATION METHOD: experimental
;   OTHER INFORMATION: /evidence= EXPERIMENTAL
;   OTHER INFORMATION: /standard_name= "Alternate Exon 3: D4.7"
;   OTHER INFORMATION: /note= "This sequence represents the third exon of
;   OTHER INFORMATION: allele D4.7 of the human D4 dopamine receptor
;   OTHER INFORMATION: gene"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 257..262
;   IDENTIFICATION METHOD: experimental
;   OTHER INFORMATION: /function= "PstI site"
;   OTHER INFORMATION: /evidence= EXPERIMENTAL
;   OTHER INFORMATION: /standard_name= "PstI site"
;   OTHER INFORMATION: /label= PstI
;   OTHER INFORMATION: /note= "This sequence is a PstI site whereby
;   OTHER INFORMATION: digestion of human genomic DNA produces a RFLP"
;   FEATURE:
;   NAME/KEY: repeat_region
;   LOCATION: 346..682
;   IDENTIFICATION METHOD: experimental
;   OTHER INFORMATION: /rpt_type= "tandem"
;   OTHER INFORMATION: /evidence= EXPERIMENTAL
;   OTHER INFORMATION: /rpt_unit= 346 .. 394
;   OTHER INFORMATION: /note= "This sequence is a repeat found in 7 known
;   OTHER INFORMATION: alleles of the human D4 dopamine receptor gene
;   OTHER INFORMATION: encoding a 16 amino acid sequence repeated 7 times
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 2..803
US-09-060-694-12

Query Match 9.2%; Score 46; DB 3; Length 803;
Best Local Similarity 49.7%; Pred. No. 0.15;
Matches 145; Conservative 0; Mismatches 145; Indels 2; Gaps 1;

```
Qy        117 GCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCC 176
Db        297 GCCGCGCGCCCCGCCGACCCAGCGGCCCTGGCCCGCCTTCCCCCACGCCACCCGCGCCCC 356

Qy        177 TCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGG 236
Db        357 GCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCCGCGCCCGGCCTTCCCCGGG 416

Qy        237 CTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296
Db        417 GTCCCTGCGGCCCCGACTGTGCGCCCGCCGCGCCCGGCCTCCCCCGGACCCCTGCGGCC 476

Qy        297 GCGCTTCTCCCGCGCCCTCAT--GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354
Db        477 CCGACTGTGCGCCCCCCGCGCCCGGCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGC 536

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCC 406
Db        537 CCCCCGCGCCCGGCCTTCCCCGGGGTCCCTGCGGCCCCGACTGTGCGCCCCC 588
```

Search completed: October 24, 2003, 15:42:52
Job time : 37.3542 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 123.415 Seconds
(without alignments)
10886.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_10000_10500
Perfect score: 501
Sequence: 1 ggttcccgtcggtctgcggg.....ggggctgggggagggggcggc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	50.8	10.1	500	10	US-09-895-040A-23	Sequence 23, Appl
	2	48.2	9.6	956	13	US-10-027-632-31508	Sequence 31508, A
c	3	47.4	9.5	1117	12	US-10-017-161-1403	Sequence 1403, Ap
c	4	47	9.4	594	12	US-10-140-472-10	Sequence 10, Appl
c	5	47	9.4	594	12	US-10-141-761-10	Sequence 10, Appl
c	6	47	9.4	594	12	US-10-142-885-10	Sequence 10, Appl
c	7	47	9.4	594	12	US-10-158-790-10	Sequence 10, Appl
c	8	47	9.4	594	14	US-10-123-155-10	Sequence 10, Appl
c	9	47	9.4	594	15	US-10-146-731-10	Sequence 10, Appl
c	10	47	9.4	3133	12	US-10-017-161-1483	Sequence 1483, Ap
	11	46.6	9.3	31168	10	US-09-764-868-1464	Sequence 1464, Ap
	12	46.4	9.3	1719	14	US-10-156-761-5084	Sequence 5084, Ap
	13	46.4	9.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
	14	46	9.2	803	14	US-10-224-260-12	Sequence 12, Appl
	15	46	9.2	1504	14	US-10-225-567A-105	Sequence 105, App
	16	46	9.2	1610	14	US-10-224-260-21	Sequence 21, Appl
c	17	45.2	9.0	2364	14	US-10-156-761-3756	Sequence 3756, Ap
	18	44.6	8.9	1074	14	US-10-156-761-5859	Sequence 5859, Ap
c	19	44.4	8.9	3163	12	US-10-017-161-1857	Sequence 1857, Ap
c	20	44.4	8.9	5452	12	US-10-017-161-1481	Sequence 1481, Ap
	21	44.2	8.8	154746	12	US-09-827-688-8	Sequence 8, Appli
c	22	44.2	8.8	154746	12	US-09-827-688-8	Sequence 8, Appli
c	23	43.8	8.7	732	12	US-10-259-165-115	Sequence 115, App
	24	43.6	8.7	3059	12	US-10-007-926A-408	Sequence 408, App
	25	43.4	8.7	1350	14	US-10-156-761-3627	Sequence 3627, Ap
	26	43.4	8.7	1614	10	US-09-976-740-45	Sequence 45, Appl
	27	43.4	8.7	1614	13	US-10-023-529-45	Sequence 45, Appl
	28	43.4	8.7	1614	13	US-10-023-523-45	Sequence 45, Appl
	29	43.4	8.7	12425	10	US-09-976-740-50	Sequence 50, Appl
	30	43.4	8.7	12425	13	US-10-023-529-50	Sequence 50, Appl
	31	43.4	8.7	12425	13	US-10-023-523-50	Sequence 50, Appl
	32	43.4	8.7	88421	10	US-09-976-059-1	Sequence 1, Appli
c	33	43.4	8.7	9025608	14	US-10-156-761-1	Sequence 1, Appli
	34	43	8.6	2307	11	US-09-883-519A-87	Sequence 87, Appl
	35	43	8.6	3670	14	US-10-074-475-98	Sequence 98, Appl
	36	42.8	8.5	1951	10	US-09-736-968A-104	Sequence 104, App
c	37	42.8	8.5	7771	10	US-09-832-292-38	Sequence 38, Appl
	38	42.6	8.5	1254	12	US-10-393-590-38	Sequence 38, Appl
	39	42.6	8.5	1254	12	US-10-393-567-38	Sequence 38, Appl
	40	42.6	8.5	1254	12	US-10-394-087-38	Sequence 38, Appl
	41	42.6	8.5	1638	14	US-10-171-581-199	Sequence 199, App
c	42	42.4	8.5	1199	12	US-10-029-386-25427	Sequence 25427, A
	43	42.4	8.5	1866	11	US-09-945-527-31	Sequence 31, Appl
	44	42.4	8.5	2124	10	US-09-815-925-8	Sequence 8, Appli
	45	42.2	8.4	1788	13	US-10-027-632-97688	Sequence 97688, A

ALIGNMENTS

RESULT 1
US-09-895-040A-23
; Sequence 23, Application US/09895040A
; Patent No. US20020123474A1
; GENERAL INFORMATION:
; APPLICANT: Shannon, Mark
; APPLICANT: Ji, Yonggang
; TITLE OF INVENTION: HUMAN GTP-RHO BINDING PROTEIN 2
; FILE REFERENCE: AEOMICA-11
; CURRENT APPLICATION NUMBER: US/09/895,040A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 09/864,761
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: Aeomica Sequence Listing Engine
; SEQ ID NO 23
; LENGTH: 500
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-895-040A-23

Query Match 10.1%; Score 50.8; DB 10; Length 500;
Best Local Similarity 45.5%; Pred. No. 0.00071;

```
  Matches  181;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy          2 GTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGA 61
Db         68 GTGCTCGGCCTCAGGCGGGGCTAGAAGGGCCGCGGGACGGGGTGGAGTGGAGGGGCGGG 127

Qy         62 ATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTC 121
Db        128 GAAGGGCGGGGACAGGGGCGGGGCCGCACGTCCTCTCGGGCCAGCCTCAGCCGCCGCGCC 187

Qy        122 CCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGC 181
Db        188 TCAGTCCGCCGTCCGCCCTCCGCGCCCGCGCCGCTAGCATGACCGACGCGCTGTTGCCCG 247

Qy        182 CCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCG 241
Db        248 CGGCCCCCCAGCCGCTGGAGAAGGAGAACGACGGCTACTTTCGGAAGGTGGGCGGCTTCG 307

Qy        242 CGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCT 301
Db        308 GGCGGGCGGGCGGGGACCTGCGGGCTCCAGACCTCCTTTCCCCCGGCCCTGCAGGGGCC 367

Qy        302 TCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGG 361
Db        368 GTGGGAGCGCGCAGTCGGGGTCCACTAGGCCGGGAGGGGGAGGGGGCGCACTGGGCCGGC 427

Qy        362 CCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAG 399
Db        428 GCTGGCCGGACGGAGGCTGGCGGGGAGGAGTGGGGGCG 465
```

RESULT 2
US-10-027-632-31508
; Sequence 31508, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 106827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 31508
;   LENGTH: 956
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(956)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-31508

```
  Query Match             9.6%;  Score 48.2;  DB 13;  Length 956;
  Best Local Similarity   46.2%;  Pred. No. 0.0033;
  Matches  146;  Conservative    1;  Mismatches  169;  Indels    0;  Gaps    0;

Qy         91 GGCTCCGGCGGCGCCAGCAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCC 150
Db        222 GGCCCCTGCCCCTGCCCCTGCCCCGCCCCCSCCCCCACCCCCCCCCCGGCCCCCCCCCC 281

Qy        151 GCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTAC 210
Db        282 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 341

Qy        211 TATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTG 270
Db        342 CCCCCCCCCCCCCCCCGGCCCCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCC 401

Qy        271 GGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGG 330
Db        402 CCCCCCCCCCCCCCCCCCCCCCTTCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 461

Qy        331 AGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGC 390
Db        462 CCCCNCCGCCCCCCCCCCCCCCCCCCCGGNCCCGGCCCCGGCNGGCCAGCGCCGCACC 521

Qy        391 GAGGCCGAGCGCGTCC 406
Db        522 CGCGGCGGCCCCCTGC 537
```

RESULT 3
US-10-017-161-1403/c
; Sequence 1403, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1403
;   LENGTH: 1117
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(1117)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(917)
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (170)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (173)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (207)..(306)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (343)..(344)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (351)..(356)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (360)..(361)

; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (378)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (380)..(383)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (398)..(399)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (401)..(402)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (405)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (431)..(432)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (434)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (437)..(438)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (440)..(444)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (464)..(465)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (468)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (484)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (487)..(497)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (509)..(510)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (525)..(527)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (532)..(533)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (539)..(542)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (548)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (552)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (557)..(561)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (565)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (573)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (577)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (588)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (605)..(610)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (615)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (617)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (619)..(622)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (629)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (646)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (648)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (650)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (652)..(653)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (669)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (673)..(675)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (688)..(689)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:

```
;   NAME/KEY: modified_base
;   LOCATION: (694)..(696)
;   OTHER INFORMATION: a, t, c, g, unknown or other
;   FEATURE:
;   NAME/KEY: modified_base
;   LOCATION: (698)

  Query Match             9.5%;  Score 47.4;  DB 12;  Length 1117;
  Best Local Similarity   37.9%;  Pred. No. 0.0053;
  Matches  108;  Conservative    0;  Mismatches  177;  Indels    0;  Gaps    0;

Qy         93 CTCCGGCGGCGCCAGCAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGC 152
Db        720 CNCNCNCCNCCCCCCCCCCCCCNCNNNCCCCNNCCCCCCCCCCCCNNNCCCNCCCCCCCC 661

Qy        153 CCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTA 212
Db        660 CCCCCCCNNCNCNCNCCCCCCCCCCCCCCCCNCCCCCCNNNNCNCNCCCCNNNNNNCCCC 601

Qy        213 TTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGG 272
Db        600 CCCCCCCCCCCCNCCCCCCCCCCCNCCCNCCCCCCCNCCCNNNNNCCCNCCCNCCCCCNN 541

Qy        273 GCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAG 332
Db        540 NNCCCCCNNCCCCNNNCCCCCCCCCCCCCCCCNNCCCCCCCCCCNNNNNNNNNNNNCCNCCC 481

Qy        333 CTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCC 377
Db        480 CCCCCCCCCCCCNCCNNCCCCCCCCCCCCCCCCCCCNNNNNCNNC 436


RESULT 4
US-10-140-472-10/c
; Sequence 10, Application US/10140472
; Publication No. US20030138888A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C168
;  CURRENT APPLICATION NUMBER: US/10/140,472
;  CURRENT FILING DATE:  2002-05-06
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-140-472-10

  Query Match             9.4%;  Score 47;  DB 12;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C----RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
Db        270 .YYSYYCSR 262


RESULT 5
US-10-141-761-10/c
; Sequence 10, Application US/10141761
; Publication No. US20030148432A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C198
;  CURRENT APPLICATION NUMBER: US/10/141,761
;  CURRENT FILING DATE:  2002-05-08
;  Prior Application removed - See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-141-761-10

  Query Match             9.4%;  Score 47;  DB 12;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C----RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391
```

```
Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
Db        270 .YYSYYCSR 262


RESULT 6
US-10-142-885-10/c
; Sequence 10, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C248
;  CURRENT APPLICATION NUMBER: US/10/142,885
;  CURRENT FILING DATE:  2002-05-10
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-142-885-10

  Query Match             9.4%;  Score 47;  DB 12;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C---- RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
Db        270 .YYSYYCSR 262


RESULT 7
US-10-158-790-10/c
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C448
;  CURRENT APPLICATION NUMBER: US/10/158,790
;  CURRENT FILING DATE:  2002-05-30
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-158-790-10

  Query Match             9.4%;  Score 47;  DB 12;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGG 167
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C--- RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
Db        270 .YYSYYCSR 262


RESULT 8
US-10-123-155-10/c
; Sequence 10, Application US/10123155
; Publication No. US20030068794A1
```

```
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C30
;  CURRENT APPLICATION NUMBER: US/10/123,155
;  CURRENT FILING DATE:  2002-04-15
; Prior Application removed -  See Palm or File Wrapper
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-123-155-10

  Query Match             9.4%;  Score 47;  DB 14;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCCCGG 167
              !  :|     :: : |   : | :  : :    :  ::::::::: : : ::::::  :
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              ::: : :  ::::::::::  :| :   : ::: ::: : |     | ::| ::::::
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C----RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              ::   :::: ::::: ::::::: ::  ::::: : :: :   :::: : :::   :
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              : ::|::: :::: ::: :  :     |       : :|: :          ::|  :
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
              |  :  :|   ::::: : ::: : :::: ::  :  : :: :  : ::  :     :
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
              :::::::
Db        270 .YYSYYCSR 262


RESULT 9
US-10-146-731-10/c
; Sequence 10, Application US/10146731
; Publication No. US20030129692A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT:  Beresini,Maureen
;  APPLICANT:  DeForge,Laura
;  APPLICANT:  Desnoyers,Luc
;  APPLICANT:  Filvaroff,Ellen
;  APPLICANT:  Gao,Wei-Qiang
;  APPLICANT:  Gerritsen,Mary E.
;  APPLICANT:  Goddard,Audrey
;  APPLICANT:  Godowski,Paul J.
;  APPLICANT:  Gurney,Austin L.
;  APPLICANT:  Sherwood,Steven
;  APPLICANT:  Smith,Victoria
;  APPLICANT:  Stewart,Timothy A.
;  APPLICANT:  Tumas,Daniel
;  APPLICANT:  Watanabe,Colin K
;  APPLICANT:  Wood,William
;  APPLICANT:  Zhang, Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION:  ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3330R1C323
;  CURRENT APPLICATION NUMBER: US/10/146,731
;  CURRENT FILING DATE:  2002-05-15
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
;   LENGTH: 594
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-146-731-10

  Query Match             9.4%;  Score 47;  DB 15;  Length 594;
  Best Local Similarity   6.8%;  Pred. No. 0.0067;
  Matches   21;  Conservative  152;  Mismatches  132;  Indels    4;  Gaps    1;

Qy        108 CAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCCCGG 167
              :  ::     :: : |   : | :  : :    :  ::::::::: : : ::::::  :
Db        566 CA.MCTT..AMM.M.CBT.STT.YA.M.YT.S.S.S.SYSYSYSYS.S.S.SDSYSYA.S 507

Qy        168 GCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCT 227
              ::: : :  ::::::::::  :| :   : ::: ::: : |     | ::| ::::::
Db        506 YSYS.S.SWSSYSYSSSDDY.CYCCYYRYHCSDSYSYSYYY.C----RCCYYT.SYSRYD 451

Qy        228 GTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTG 287
              ::   :::: ::::: ::::::: ::  ::::: : :: :   :::: : :::   :
Db        450 CHYSCCCSDYYCYSYSYSRYYSYSYSYSWSYSYSYTDYCSYRRCCCYYSYSSSYSSSYSS 391

Qy        288 GCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCC 347
              : ::|::: :::: ::: :  :     |       : :|: :          ::|  :
Db        390 SAYSTSSSSSSSSSYYYTSTNYC.T.CC....T..MCAABCSTTTTTTTTTT..HSCC.S 331

Qy        348 AGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCG 407
              |  :  :|   ::::: : ::: : :::: ::  :  : :: :  : ::  :     :
Db        330 A..A.M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

Qy        408 AGTCTTCCA 416
              :::::::
Db        270 .YYSYYCSR 262


RESULT 10
US-10-017-161-1483/c
; Sequence 1483, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1483
;   LENGTH: 3133
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
```

; LOCATION: source
; FEATURE:
; LOCATION: (1)..(3133)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(2933)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1)..(118)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (124)..(129)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (146)..(148)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (151)..(152)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (155)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (172)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (179)..(180)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (182)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (185)..(188)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (192)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (194)..(195)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (210)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (212)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (214)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (219)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (222)..(225)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (229)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (231)..(232)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (238)..(239)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (242)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (245)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (248)..(260)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (265)..(270)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (272)..(276)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (278)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (282)..(283)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (287)..(290)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (302)..(306)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (311)..(313)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (323)..(334)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (336)..(341)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (345)..(363)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (365)..(368)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370)
; OTHER INFORMATION: a, t, c, g, unknown or other

```
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (374)..(375)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (379)..(385)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (387)..(388)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (390)..(393)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (395)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (397)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (400)..(402)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base
;  LOCATION: (409)
;  OTHER INFORMATION: a, t, c, g, unknown or other
;  FEATURE:
;  NAME/KEY: modified_base

  Query Match             9.4%;  Score 47;  DB 12;  Length 3133;
  Best Local Similarity   37.9%;  Pred. No. 0.0068;
  Matches  107;  Conservative    0;  Mismatches  175;  Indels    0;  Gaps    0;

Qy        102 CGCCAGCAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCGCCCCTCCCGC 161

Db        721 CCCCCCCCCCCCNCCNCNCCCCNNCCCCCCCCCCCCCCCNNNNCCCCCCCCNNCCCCCC 662

Qy        162 CGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTA 221

Db        661 CCCCCCNCCNNNNNCCCCNCCCCCNNCCCCCCCCCCCCCNNNCCNCNCCCCCCCCCCCNN 602

Qy        222 CCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTT 281

Db        601 CCCCCCNNCNNCCCCCCCCCCCCNNNCCCNCCCCNNNNCCCCCCCCCCCNNNNNCCCCCN 542

Qy        282 CGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGC 341

Db        541 CCCCCNCCNCCCCCCCCCNNNNCCCCCCCCCCCCCNNCCCNCCCNNCCCNNNNCCCCNNC 482

Qy        342 CGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCC 383

Db        481 CCCNCCCCNCCCCCCCCCCNNCNCCCCNCCNNNCNNNNCCCCC 440

RESULT 11
US-09-764-868-1464
; Sequence 1464, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PZ032
;  CURRENT APPLICATION NUMBER: US/09/764,868
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - refer to PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 1510
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1464
;   LENGTH: 31168
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-868-1464

  Query Match             9.3%;  Score 46.6;  DB 10;  Length 31168;
  Best Local Similarity   55.2%;  Pred. No. 0.0087;
  Matches   91;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        337 GGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCC 396

Db      17294 GGGTCCTGGCCAGGGCCAGCCGCAGCATCCACACCCACCCCACAGCTGGATGGCTCCTTG 17353

Qy        397 GAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATC 456

Db      17354 GAGGACAACCTAGACCTCCCTCACCACGGCCTGGCCGACAGCTCCGCTGACCTCAGCTGC 17413

Qy        457 GATGAGGATGAGAAAGGTAACTAGGGGGCTGGGGAGGGGGCGGC 501

Db      17414 GACAGCCGCGTTAACAGCGACTACGAGACGGACGGCGAGGGCGGC 17458

RESULT 12
US-10-156-761-5084
; Sequence 5084, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5084
;   LENGTH: 1719
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1719)
US-10-156-761-5084

  Query Match             9.3%;  Score 46.4;  DB 14;  Length 1719;
  Best Local Similarity   52.0%;  Pred. No. 0.0097;
  Matches  104;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        257 TCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCA 316

Db        806 TCGCGCGCGCACTGCGCGTCGTCGCCGGGGGCGTCGGCGTGCACGCGAGTGCCGCGCAGA 865

Qy        317 TGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGC 376

Db        866 TCGCCGCCGCGGAGGGCGTCGGGGCGCTGCTGGCCCCCGATCCGACACCCGCCACGGTGC 925

Qy        377 CGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACA 436

Db        926 CCGGCGTGCCCGGCGCGGCCGACCCCACGCAGGTGCTGCCCCAGGGCGCGGGCGCGTACG 985

Qy        437 TCTCCATTGCCCTGCGCATC 456

Db        986 ACCCGAACGCGGCGACCAGC 1005
```

```
RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
;   LENGTH: 9025608
;   TYPE: DNA
;   ORGANISM: Streptomyces avermitilis
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (4187715)
;   OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

  Query Match             9.3%;  Score 46.4;  DB 14;  Length 9025608;
  Best Local Similarity   52.0%;  Pred. No. 0.01;
  Matches  104;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy        257 TCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCA 316

Db    6193898 TCGCGCGCGCACTGCGCGTCGTCGCCGGGGGCGTCGGCGTGCACGCGAGTGCCGCGCAGA 6193957

Qy        317 TGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGC 376

Db    6193958 TCGCCGCCGCGGAGGGCGTCGGGGCGCTGCTGGCCCCCGATCCGACACCCGCCACGGTGC 6194017

Qy        377 CGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACA 436

Db    6194018 CCGGCGTGCCCGGCGCGGCCGACCCACGCAGGTGCTGCCCCAGGGCGCGGGCGCGTACG 6194077

Qy        437 TCTCCATTGCCCTGCGCATC 456

Db    6194078 ACCCGAACGCGGCGACCAGC 6194097


RESULT 14
US-10-224-260-12
; Sequence 12, Application US/10224260
; Publication No. US20030059845A1
;   GENERAL INFORMATION:
;        APPLICANT: Van Tol, Hubert H.M.
;                   Civelli, Olivier
;        TITLE OF INVENTION: A No. US20030059845Alel Human Dopamine Receptor and Uses
;        NUMBER OF SEQUENCES: 22
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Allegretti & Witcoff, Ltd.
;             STREET: 10 South Wacker Drive, Suite 3000
;             CITY: Chicago
;             STATE: Illinois
;             COUNTRY: USA
;             ZIP: 60606
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.25
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/224,260
;             FILING DATE: 20-Aug-2002
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 07/928,611
;             FILING DATE: <Unknown>
;        ATTORNEY/AGENT INFORMATION:
;             NAME: No. US20030059845Alnan, Kevin E
;             REGISTRATION NUMBER: 35,303
;             REFERENCE/DOCKET NUMBER: 90,1092-B
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 312-715-1000
;             TELEFAX: 312-715-1234
;             TELEX: 910-221-8317
;   INFORMATION FOR SEQ ID NO: 12:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 803 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        FEATURE:
;             NAME/KEY:  exon
;             LOCATION:  1..803
;             IDENTIFICATION METHOD:  experimental
;             OTHER INFORMATION:  /evidence= EXPERIMENTAL
;             /standard_name= "Alternate Exon 3: D4.7"
;             /note= "This sequence represents the third exon of
;             allele D4.7 of the human D4 dopamine receptor
;             gene"
;        FEATURE:
;             NAME/KEY:  misc_feature
;             LOCATION:  257..262
;             IDENTIFICATION METHOD:  experimental
;             OTHER INFORMATION:  /function= "PstI site"
;             /evidence= EXPERIMENTAL
;             /standard_name= "PstI site"
;             /label= PstI
;             /note= "This sequence is a PstI site whereby
;             digestion of human genomic DNA produces a RFLP"
;        FEATURE:
;             NAME/KEY:  repeat_region
;             LOCATION:  346..682
;             IDENTIFICATION METHOD:  experimental
;             OTHER INFORMATION:  /rpt_type= "tandem"
;             /evidence= EXPERIMENTAL
;             /rpt_unit= 346 .. 394
;             /note= "This sequence is a repeat found in 7 known
;             alleles of the human D4 dopamine receptor gene
;             encoding a 16 amino acid sequence repeated 7 times
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  2..803
;        SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-224-260-12

  Query Match             9.2%;  Score 46;  DB 14;  Length 803;
  Best Local Similarity   49.7%;  Pred. No. 0.012;
  Matches  145;  Conservative    0;  Mismatches  145;  Indels    2;  Gaps    1;

Qy        117 GCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCC 176

Db        297 GCCGCGCGCCCCGCCGACCCAGCGGCCCTGGCCCGCCTTCCCCCACGCCACCCGCGCCCC 356

Qy        177 TCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGG 236

Db        357 GCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCCGCGCCCGGCCTTCCCCGGG 416

Qy        237 CTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296

Db        417 GTCCCTGCGGCCCCGACTGTGCGCCCGCCGCGCCCGGCCTCCCCCCGGACCCCTGCGGCC 476
```

```
Qy        297 GCGCTTCTCCCGCGCCCTCAT--GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354

Db        477 CCGACTGTGCGCCCCCCGCGCCCGGCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGC 536

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCC 406

Db        537 CCCCCGCGCCCGGCCTTCCCCGGGGTCCCTGCGGCCCCGACTGTGCGCCCCC 588


RESULT 15
US-10-225-567A-105
; Sequence 105, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
;  APPLICANT: LifeSpan Biosciences
;  APPLICANT:  Brown, Joseph P.
;  APPLICANT:  Burmer, Glenna C.
;  APPLICANT:  Roush, Christine L.
;  TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 1920-4-4
;  CURRENT APPLICATION NUMBER: US/10/225,567A
;  CURRENT FILING DATE:  2001-12-19
;  PRIOR APPLICATION NUMBER: 60/257,144
;  PRIOR FILING DATE: 2000-12-19
;  NUMBER OF SEQ ID NOS: 2292
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 105
;   LENGTH: 1504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-225-567A-105

  Query Match             9.2%;  Score 46;  DB 14;  Length 1504;
  Best Local Similarity   49.7%;  Pred. No. 0.012;
  Matches  145;  Conservative    0;  Mismatches  145;  Indels    2;  Gaps    1;

Qy        117 GCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCC 176

Db        695 GCCGCGCGCCCCGCCGACCCAGCGGCCCTGGCCCGCCTTCCCCCACGCCACCCGCGCCCC 754

Qy        177 TCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGG 236

Db        755 GCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGCCCCCCGCGCCCGGCCTTCCCCGGG 814

Qy        237 CTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296

Db        815 GTCCCTGCGGCCCCGACTGTGCGCCCGCCGCGCCCGGCCTCCCCCGGACCCCTGCGGCC 874

Qy        297 GCGCTTCTCCCGCGCCCTCAT--GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354

Db        875 CCGACTGTGCGCCCCCCGCGCCCGGCCTCCCCCAGGACCCCTGCGGCCCCGACTGTGCGC 934

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCC 406

Db        935 CCCCCGCGCCCGGCCTTCCCCGGGGTCCCTGCGGCCCCGACTGTGCGCCCCC 986


Search completed: October 24, 2003, 18:34:43
Job time : 140.415 secs
```

GenCore version 5.1.6

CM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 1046.7 Seconds
(without alignments)
11633.300 Million cell updates/sec

Title: US-09-830-902-1_COPY_10000_10500
Perfect score: 501
Sequence: 1 ggtttccgtcggtctgcggg.....gggatggggagggggcggc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	344	68.7	916	13	BQ944322	BQ944322 AGENCOURT
2	281.2	56.1	1106	13	BU518568	BU518568 AGENCOURT
3	248.2	49.5	517	28	AQ937943	AQ937943 NB6-679R
4	246	49.1	2897	11	AK044900	AK044900 Mus muscu
5	239.6	47.8	693	10	BB643122	BB643122 BB643122
6	224.8	44.9	600	12	BI987009	BI987009 3184-70 M
7	224.2	44.8	557	12	BM506397	BM506397 1143d05.y
8	220.2	44.0	384	13	BY261238	BY261238 BY261238
9	216	43.1	976	10	BG391372	BG391372 602417440
10	208.8	41.7	553	10	BB618976	BB618976 BB618976
11	208.6	41.6	1843	11	AK007793	AK007793 Mus muscu
12	197.6	39.4	597	14	BY707801	BY707801 BY707801
13	196.4	39.2	375	14	CB691737	CB691737 AMGNNUC:N
14	188.2	37.6	486	28	AQ937698	AQ937698 NB6-323R
15	186.6	37.2	548	28	AQ683149	AQ683149 HS_5378_B
16	186.6	37.2	890	10	BF178630	BF178630 601806609
17	183.2	36.6	589	10	BE373511	BE373511 601225864
18	172.8	34.5	569	10	BG625233	BG625233 pgn1c.pk0
19	170.2	34.0	641	12	BM491941	BM491941 pgp2n.pk0
20	164	32.7	604	9	AJ450818	AJ450818 AJ450818
21	163.4	32.6	659	9	AJ446053	AJ446053 AJ446053
22	141.4	28.2	677	9	AL628068	AL628068 AL628068
23	141.2	28.2	662	9	AL649034	AL649034 AL649034
24	139	27.7	617	9	AL635035	AL635035 AL635035
25	138	27.5	623	9	AL655040	AL655040 AL655040
26	135.8	27.1	615	9	AL972334	AL972334 AL972334
27	128.8	25.7	372	10	BG724026	BG724026 602697189
28	125	25.0	648	9	AL873602	AL873602 AL873602
29	124	24.8	644	9	AL871548	AL871548 AL871548
30	117.8	23.5	880	13	BU907712	BU907712 AGENCOURT
31	115.8	23.1	680	11	AF549969	AF549969 Xenopus l
32	115	23.0	316	13	BU057119	BU057119 UI-M-FY0-
33	105.4	21.0	379	14	CD348493	CD348493 UI-M-FY0-
34	103	20.6	655	10	BB619871	BB619871 BB619871
35	99.8	19.9	529	14	CB018332	CB018332 pgn1c.pk0
36	91.6	18.3	365	10	BG234069	BG234069 daa36e02.
37	80	16.0	683	9	AL628360	AL628360 AL628360
38	79.4	15.8	418	9	AA560327	AA560327 v167e05.r
39	77.8	15.5	277	9	AL856941	AL856941 AL856941
40	77.8	15.5	929	13	BU916112	BU916112 AGENCOURT
41	75.8	15.1	366	12	BI348284	BI348284 dai27g03.
42	71.8	14.3	925	29	CNS0091P	AL053013 Drosophil
43	66.8	13.3	982	13	BX415111	BX415111 BX415111
c 44	66.2	13.2	1057	13	BX349688	BX349688 BX349688
c 45	64.8	12.9	982	13	BX415111	BX415111 BX415111

ALIGNMENTS

RESULT 1
BQ944322
LOCUS BQ944322 916 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_8858958 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:6441742 5', mRNA sequence.
ACCESSION BQ944322
VERSION BQ944322.1 GI:22359800
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 916)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13965 row: g column: 23
High quality sequence stop: 478.

```
FEATURES             Location/Qualifiers
     source          1. .916
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N-3"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6441742"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Mam2"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      161 a    347 c    272 g    136 t
ORIGIN

  Query Match           68.7%;  Score 344;  DB 13;  Length 916;
  Best Local Similarity   86.0%;  Pred. No. 4.3e-59;
  Matches  406;  Conservative    0;  Mismatches   60;  Indels    6;  Gaps    2;

Qy         10 CGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATGAATTCTCCG 69
Db          1 CGGTCTGCGGGAAGCGGGTTATGGCGGCGGCGGCAGTGGGAGCTGTGAATGAGTTCTCCG 60

Qy         70 GGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCTCCCAGGCCT 129
Db         61 GCCGGACGACGGAAGAAGAAAGGCTCGGGCGGCGCGAG---CCCGGCGCCCGCCAGGCCT 117

Qy        130 CCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCGCCCGAGTCG 189
Db        118 CCGCCCCCCGCCGCGGTCCCCGCCCCT---GCCGCCGGCCCGGCCCCTGCGGCCGGCTCG 174

Qy        190 CCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTG 249
Db        175 CCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTCGCCCTGCTG 234

Qy        250 CGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGC 309
Db        235 CGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGGCGTGGCTCTGCCAGCGCTTCTCCGC 294

Qy        310 GCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCC 369
Db        295 GCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCGCCCTCGCCC 354

Qy        370 CCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTC 429
Db        355 CCAGCGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCTTC 414

Qy        430 GAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGG 481
Db        415 GAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGACAGAAGG 466


RESULT 2
BU518568
LOCUS       BU518568               1106 bp    mRNA    linear   EST 12-SEP-2002
DEFINITION  AGENCOURT_10164034 NIH_MGC_134 Mus musculus cDNA clone
            IMAGE:6516310 5', mRNA sequence.
ACCESSION   BU518568
VERSION     BU518568.1  GI:22826094
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1106)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
            cDNA Library Preparation: Invitrogen Corp
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14094  row: j  column: 23
            High quality sequence start: 17
            High quality sequence stop: 520.
FEATURES             Location/Qualifiers
     source          1. .1106
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6516310"
                     /tissue_type="undifferentiated limb"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_134"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Cloned unidirectionally. Primer: Oligo dT. Average
                     insert size 1.7 kb. Constructed by ResGen, Invitrogen
                     Corp. Note: this is a NIH_MGC Library."
BASE COUNT      182 a    379 c    325 g    220 t
ORIGIN

  Query Match           56.1%;  Score 281.2;  DB 13;  Length 1106;
  Best Local Similarity   84.7%;  Pred. No. 1.6e-46;
  Matches  364;  Conservative    0;  Mismatches   58;  Indels    8;  Gaps    4;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60
Db        158 GGTCGCCGTCGGTCTGCGGGAAGCGGGTTATGGCGGCGGCGGCAGTGGGAGCTGTGAATG 217

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120
Db        218 AGTTCTCCGGCCGGACGACGGAAGAAGAAAGGCTCGGGCGGCGCGAG---CCCGGCGCCC 274

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180
Db        275 GCCAGGCCTCCGCCCCCCGCCGCGGTCCCCGCCCCT---GCCGCCGGCCCGGCCCCTGCG 331

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240
Db        332 GCCGGCTCGCCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTC 391

Qy        241 GCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGC 300
Db        392 GCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGC 451

Qy        301 TTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCG 360
Db        452 TTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCG 511

Qy        361 GCCTCGGCCCCGGCG-CCGGTGCCGGGCGGCGAGGCCGAGCGC-GTCCGAGTCTTCCACA 418
Db        512 CCCTCGCCCCCAGCGCCCGGGACCGGGTGGCGAGGCCGAAAACGGTCCGCGTCTTCCCCA 571

Qy        419 AACAGGCCTT 428
Db        572 ATCAGGCCTT 581


RESULT 3
AQ937943
LOCUS       AQ937943               517 bp    DNA     linear   GSS 23-AUG-2000
DEFINITION  NB6-679R Human NotI clones Homo sapiens genomic, genomic survey
            sequence.
ACCESSION   AQ937943
VERSION     AQ937943.1  GI:7214321
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 517)
  AUTHORS   Zabarovsky,E.R., Gizatullin,R., Podowski,R.M., Zabarovska,V.V., Xie
            ,L., Muravenko,O.V., Kozyrev,S., Petrenko,L., Skobeleva,N., Li,J.,
            Protopopov,A., Kashuba,V., Ernberg,I., Winberg,G. and Wahlestedt,C.
  TITLE     NotI clones in the analysis of the human genome
  JOURNAL   Nucleic Acids Res. 28 (7), 1635-1639 (2000)
  MEDLINE   20175728
   PUBMED   10710430
COMMENT     Contact: Podowski RM
            Center for Genomics Research
            Karolinska Institute
            17177 Stockholm, Sweden
            Tel: +46-8-729-6372
            Fax: +46-8-337983
            Email: Raf.Podowski@cgr.ki.se
            Class: NotI site.
FEATURES             Location/Qualifiers
     source          1..517
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human NotI clones"
BASE COUNT       76 a    158 c    190 g     89 t      4 others
ORIGIN

  Query Match             49.5%;  Score 248.2;  DB 28;  Length 517;
  Best Local Similarity   93.6%;  Pred. No. 6e-40;
  Matches  278;  Conservative    0;  Mismatches   17;  Indels    2;  Gaps    2;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGCGGGTTATGGCGGCGGCGGCAGTGAGAGCTGTGAATG 60

Db        221 GGTTCCCGTCGGTCTGCTGGAGGCGGGTTATGGCGGCGGCNGCAGTGAGAGCTGNGAATG 280

Qy         61 AATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 120

Db        281 AATTCTCCGGGTGGACAAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGCCT 340

Qy        121 CCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTCCG 180

Db        341 CCCAAGCCTTCGCCCTCTTGTCTGGTCCCCGCCCCTCCCGNCGGCGGGCCGGCCNCTGCG 400

Qy        181 CCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTC 240

Db        401 TCCGAGTCGTCGCATAAGCGGAACCTGTACTATTT-TCCTACCCGCTGTTTGTAGGCTTC 459

Qy        241 GCGCTGC-TGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCA 296

Db        460 GCGCTGCTTGCATTTGGTCGCCTTCCACCTGGGGCTCCTCTTTGTGTGGCTCTGCCA 516


RESULT 4
AK044900
LOCUS       AK044900                2897 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 9.5 days embryo parthenogenote cDNA, RIKEN full-length
            enriched library, clone:B130010N24 product:spastic paraplegia 4
            homolog (human), full insert sequence.
ACCESSION   AK044900
VERSION     AK044900.1  GI:26090599
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2897)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
```

```
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Tissues were provided by Dr. Tomohiro Kono (Department of Animal
          Science, Tokyo University of Agriculture, 1737 Hunako Atsugi City,
          Kanagawa Prefecture, Japan) whose assistance we gratefully
          acknowledge.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2897
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B130010N24"
                     /db_xref="taxon:10090"
                     /clone="B130010N24"
                     /tissue_type="parthenogenote"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="9.5 days embryo"
     misc_feature    1..2897
                     /note="spastic paraplegia 4 homolog (human)
                     (MGD MGI:1858896, GB AK077794, evidence: BLASTN, 100%,
                     match=1843)"
BASE COUNT      822 a    632 c    687 g    756 t
ORIGIN

  Query Match            49.1%;  Score 246;  DB 11;  Length 2897;
  Best Local Similarity  87.1%;  Pred. No. 1.8e-59;
  Matches  270;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        163 GCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTAC 222

Db          2 GCCGGCCCGGCCCCTGCGGCCGGCTCGCCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCC 61

Qy        223 CCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTC 282

Db         62 CCGCTGGTCGTCGGCTTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTC 121

Qy        283 GTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCC 342

Db        122 GCGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCGGCCAAGAGGAGCTCCGGGACC 181

Qy        343 GCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGC 402

Db        182 GCGCCGGCGCCCGCCTCGCCCTCGCCCCCAGAGCCCGGACCGGGTGGCGAGGCCGAGAGC 241

Qy        403 GTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAG 462

Db        242 GTCCGCGTCTTCCACAAGCAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGAG 301

Qy        463 GATGAGAAAG 472

Db        302 GAAGAGAAAG 311


RESULT 5
BB643122
LOCUS       BB643122                 693 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB643122 RIKEN full-length enriched, 9.5 days embryo parthenogenote
            Mus musculus cDNA clone B130010N24 5', mRNA sequence.
ACCESSION   BB643122
VERSION     BB643122.1  GI:16477873
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 693)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..693
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="B130010N24"
                     /tissue_type="parthenogenote"
                     /dev_stage="9.5 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 9.5 days embryo
                     parthenogenote"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of subtraction to
                     Rot = 229.0 Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with XhoI and BamHI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      160 a    206 c    194 g    133 t
ORIGIN

  Query Match            47.8%;  Score 239.6;  DB 10;  Length 693;
  Best Local Similarity  87.1%;  Pred. No. 3.2e-38;
  Matches  263;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;
```

```
Qy        171 GGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTT 230
Db          1 GGCCCCTGCGGCCGGCTCGCCGCCTAAGCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGT 60

Qy        231 TGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCT 290
Db         61 CGTCGGCTTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCT 120

Qy        291 CTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGC 350
Db        121 CTGCCAGCGCTTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGC 180

Qy        351 ACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGT 410
Db        181 GCCCGCCTCGCCCTCGCCCCCAGAGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGT 240

Qy        411 CTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAA 470
Db        241 CTTCCACAAGCAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAA 300

Qy        471 AG 472
Db        301 AG 302


RESULT 6
BI987009
LOCUS       BI987009                 600 bp    mRNA    linear   EST 20-DEC-2001
DEFINITION  3184-70 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,
            mRNA sequence.
ACCESSION   BI987009
VERSION     BI987009.1  GI:17957970
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 600)
  AUTHORS   Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
            White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
  TITLE     Gene expression in the developing mouse retina by EST sequencing
            and microarray analysis
  JOURNAL   Nucleic Acids Res. 29 (24), 4983-4993 (2001)
  MEDLINE   21671825
   PUBMED   11812828
COMMENT     Contact: Klein WH
            Department of Biochemistry and Molecular Biology
            University of Texas M.D. Anderson Cancer Center
            Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
            Tel: 713 792 3646
            Fax: 713 790 0329.
FEATURES             Location/Qualifiers
     source          1..600
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /tissue_type="neural retina"
                     /dev_stage="embryonic day 14.5 post-fertilization"
                     /clone_lib="Mouse E14.5 retina lambda ZAP II Library"
BASE COUNT      144 a    176 c    161 g    119 t
ORIGIN

  Query Match            44.9%;  Score 224.6;  DB 12;  Length 600;
  Best Local Similarity  87.0%;  Pred. No. 2.9e-35;
  Matches  247;  Conservative    0;  Mismatches   37;  Indels    0;  Gaps    0;

Qy        198 GCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGT 257
Db          1 GCGGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTCGCCCTGCTGCGCCTGCT 60

Qy        258 CGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCAT 317
Db         61 GGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCAT 120

Qy        318 GGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCC 377
Db        121 GGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCGCCCTCGCCCCCAGCGCC 180

Qy        378 GGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACAT 437
Db        181 CGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCTTCGAGTACAT 240

Qy        438 CTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGG 481
Db        241 CTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGACAGAAGG 284


RESULT 7
BM506397
LOCUS       BM506397                 557 bp    mRNA    linear   EST 12-MAR-2002
DEFINITION  i143d05.y1 Melton Normalized Mixed Mouse Pancreas 1 N1-MMS1 Mus
            musculus cDNA clone IMAGE:5944376 5' similar to TR:Q9UPR9 Q9UPR9
            KIAA1083 PROTEIN. [1] ;, mRNA sequence.
ACCESSION   BM506397
VERSION     BM506397.1  GI:18670439
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 557)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished
COMMENT     Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1912
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Douglas Melton DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Juliana Brown
            (brown@fas.harvard.edu)
            MGI:2007249 This sequence now available from the IMAGE consortium,
            for clone orders contact: info@image.llnl.gov
            Seq primer: -40RP from Gibco
            High quality sequence stop: 435.
FEATURES             Location/Qualifiers
     source          1..557
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="ICR"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5944376"
                     /sex="Both for embryonic & newborn, male for adult and
                     adult islet"
                     /dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
                     adult, mixed"
                     /lab_host="DH10B"
                     /clone_lib="Melton Normalized Mixed Mouse Pancreas 1
                     N1-MMS1"
                     /note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; Five
                     libraries representing E10.5/12.5 pancreatic bud, E16.5
                     pancreas,  newborn pancreas, adult pancreas, and adult
                     islets of Langerhans were  seperately constructed using
                     SuperScript Plasmid Library kit (Life  Technologies). cDNA
                     was made by oligo-dT priming and size-selected by  column
                     fractionation. Libraries were amplified once on solid
```

```
            support  and plasmid DNA from each library was prepared
            and mixed in equal  amounts. The mixed library DNA was
            normalized by method #4 from  Bonaldo, Lennon, and Soares
            1996 Genome Research 6:791-806; 0.5  microgram
            single-stranded mixed library plasmid DNA was mixed with
            5 micrograms PCR product representing mixed library
            inserts and  hybridized to an Ecot of 6. Single-stranded
            (unhybridized) plasmids  were isolated by hydroxyapatite
            chromatography and used to make this  library."
BASE COUNT      132 a    164 c    152 g    109 t
ORIGIN

  Query Match             44.8%;  Score 224.2;  DB 12;  Length 557;
  Best Local Similarity   88.1%;  Pred. No. 3.8e-35;
  Matches  244;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy        196 AAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTG 255
Db          1 AAGCGGAACCTTTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTCGCCCTGCTGCGCCTG 60

Qy        256 GTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTC 315
Db         61 CTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTC 120

Qy        316 ATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCG 375
Db        121 ATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCGCCCTCGCCCCCAGCG 180

Qy        376 CCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTAC 435
Db        181 CCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCACAAGCAGGCCTTCGAGTAC 240

Qy        436 ATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
Db        241 ATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAG 277


RESULT 8
BY261238
LOCUS       BY261238                 384 bp    mRNA    linear   EST 10-DEC-2002
DEFINITION  BY261238 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K330305P12 5', mRNA sequence.
ACCESSION   BY261238
VERSION     BY261238.1  GI:26442750
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 384)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Tissues were provided by Michela Fagiolini and Takao K. Hensch (
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge. Please visit our web site
            (http://genome.gsc.riken.go.jp) for further details.
FEATURES             Location/Qualifiers
     source          1..384
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K330305P12"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT       79 a    123 c    116 g     65 t      1 others
ORIGIN

  Query Match             44.0%;  Score 220.2;  DB 13;  Length 384;
  Best Local Similarity   86.2%;  Pred. No. 2.3e-34;
  Matches  243;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy        200 GGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCG 259
Db          2 GGAACCCGTCTTCTTTCTCGTCCCCGCTGGTCGTCGGCTTCGCCCTGCTGCGCCTGCTGG 61

Qy        260 CCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGG 319
Db         62 CCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGG 121

Qy        320 CAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGG 379
Db        122 CCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCCTCGCCCTCGCCCCCAGAGCCCG 181

Qy        380 TGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCT 439
```

```
Db        182 GACCGGGTGGCGAGGCCGAGAGCGTCCGNGTCTTCCACAAGCAGGCCTTCGAGTACATAT 241

Qy        440 CCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAACTAGG 481

Db        242 CCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGACAGAAGG 283


RESULT 9
BG391372
LOCUS       BG391372                 976 bp    mRNA    linear   EST 12-MAR-2001
DEFINITION  602417440F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4537184 5',
            mRNA sequence.
ACCESSION   BG391372
VERSION     BG391372.1  GI:13284830
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10461  row: o  column: 09
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..976
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4537184"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_92"
                     /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.3 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
BASE COUNT      287 a    237 c    253 g    199 t
ORIGIN

  Query Match             43.1%;  Score 216;  DB 10;  Length 976;
  Best Local Similarity   99.2%;  Pred. No. 1.7e-33;
  Matches  238;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy        233 TAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCT 292

Db          1 TAGGCTTCGCGCTGCTGCG-TTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCT 59

Qy        293 GCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCAC 352

Db         60 GCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCAC 119

Qy        353 CTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCT 412

Db        120 CTGCCTCGGCCTCGG-CCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCT 178

Qy        413 TCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472

Db        179 TCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 238
```

```
RESULT 10
BB618976
LOCUS       BB618976                 553 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB618976 RIKEN full-length enriched, 8 days embryo Mus musculus
            cDNA clone 5730458M19 5', mRNA sequence.
ACCESSION   BB618976
VERSION     BB618976.1  GI:16458433
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
             e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..553
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="5730458M19"
                     /sex="mixed"
                     /dev_stage="8 days embryo"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 8 days embryo"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
```

primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 137 a 160 c 152 g 104 t
ORIGIN

Query Match 41.7%; Score 208.8; DB 10; Length 553;
Best Local Similarity 87.7%; Pred. No. 4.5e-32;
Matches 228; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy      222 CCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTT 281
Db        1 CCCGCTGGTCGTCGGCTTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTT 60

Qy      282 CGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGC 341
Db       61 CGCGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGAC 120

Qy      342 CGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCG 401
Db      121 CGCGCCGGCGCCCGCCTCGCCCTCGCCCCAGAGCCCGGACCGGGTGGCGAGGCCGAGAG 180

Qy      402 CGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGA 461
Db      181 CGTCCGCGTCTTCCACAAGCAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGA 240

Qy      462 GGATGAGAAAGGTAACTAGG 481
Db      241 GGAAGAGAAAGGACAGAAGG 260
```

RESULT 11
AK007793
LOCUS AK007793 1843 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810046H01 product:spastic paraplegia 4 homolog (human), full insert sequence.
ACCESSION AK007793
VERSION AK007793.1 GI:12841565
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system- 384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1843)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAGCGGCCGCAATTAATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES Location/Qualifiers

```
     source          1..1843
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:1810046H01"
                     /db_xref="MGI:1934768"
                     /db_xref="taxon:10090"
                     /clone="1810046H01"
                     /sex="male"
                     /tissue_type="pancreas"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 day old"
     CDS             <1..1672
                     /note="unnamed protein product; putative
                     spastic paraplegia 4 homolog (human) (MGD|MGI:1858896,
                     GB|AK027793, evidence: BLASTN, 100%, match=1843)"
                     /codon_start=2
                     /protein_id="BAB25259.1"
                     /db_xref="GI:12841566"
                     /db_xref="MGI:1858896"
                     /translation="GPALLRLLACHLGLLFAWLCQRFSRALMAAKRSSGTAPAPASPS
                     PPEPGPGGEAESVRVFHKQAFEYISIALRIDEEEKGQKEQAVEWYKKGIEELEKGIAV
                     IVTGQGEQYERARRLQAKMMTNLVMAKDRLQLLEKLQPVLQFSKSQTDVYNESTNLTC
                     RNGHLQSESGAVPKRKDPLTHASNSLPRSKTVLKSGSAGLSGHHRAPSCSGLSMVSGA
                     RPGPGPAATTHKGTPKPNRTNKPSTPTTAVRKKKDLKNFRNVDSNLANLIMNEIVDNG
                     TAVKFDDIAGQELAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLAKA
                     VAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCERRE
                     GEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSLPN
                     EETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKPEQ
                     VKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     polyA_signal    1822..1827
                     /note="putative"
     polyA_site      1843
                     /note="putative"
BASE COUNT      535 a    427 c    466 g    415 t
ORIGIN

  Query Match          40.0%;  Score 200.6;  DB 11;  Length 1843;
  Best Local Similarity   88.3%;  Pred. No. 2.1e-30;
  Matches  218;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        235 GGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGC 294
              
Db          2 GGCTTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGC 61

Qy        295 CAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354
              
Db         62 CAGCGCTTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCC 121

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTC 414
              
Db        122 GCCTCGCCCTCGCCCCCAGAGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTC 181

Qy        415 CACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGT 474
              
Db        182 CACAAGCAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGA 241

Qy        475 AACTAGG 481
              
Db        242 CAGAAGG 248


RESULT 12
BY707801
LOCUS       BY707801                 597 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY707801 RIKEN full-length enriched, 10 day old male pancreas Mus
            musculus cDNA clone 1810046H01 5', mRNA sequence.
ACCESSION   BY707801
VERSION     BY707801.1  GI:27118981
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..597
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
```

```
                  /db_xref="taxon:10090"
                  /clone="1810046H01"
                  /sex="male"
                  /tissue_type="pancreas"
                  /dev_stage="10 day old"
                  /clone_lib="RIKEN full-length enriched, 10 day old male
                  pancreas"
BASE COUNT      154 a    169 c    159 g    115 t
ORIGIN

  Query Match             39.4%;  Score 197.6;  DB 14;  Length 597;
  Best Local Similarity   88.1%;  Pred. No. 7.9e-30;
  Matches  215;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy        238 TTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGCCAG 297
Db          1 TTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGCGTGGCTCTGCCAG 60

Qy        298 CGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCTGCC 357
Db         61 CGCTTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCCGCC 120

Qy        358 TCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTCCAC 417
Db        121 TCGCCCTCGCCCCCAGAGCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTCCAC 180

Qy        418 AAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAGGTAAC 477
Db        181 AAGCAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAGGACAG 240

Qy        478 TAGG 481
Db        241 AAGG 244


RESULT 13
CB691737
LOCUS       CB691737                 375 bp    mRNA    linear   EST 10-APR-2003
DEFINITION  AMGNNUC:NRDG1-00091-G1-A nrdg1 (10855) Rattus norvegicus cDNA clone
            nrdg1-00091-g1 5', mRNA sequence.
ACCESSION   CB691737
VERSION     CB691737.1  GI:29749884
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Amgen EST Program.
  TITLE     Amgen Rat EST Program
  JOURNAL   Unpublished
COMMENT     Contact: Dan Fitzpatrick
            Amgen, Inc
            One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
            Tel: 805 447-4881
            Plate: 00091  row: g  column: 1.
FEATURES             Location/Qualifiers
     source          1..375
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
                     /clone="nrdg1-00091-g1"
                     /tissue_type="Dorsal Root Ganglia"
                     /clone_lib="nrdg1 (10855)"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; rat
                     dorsal root ganglia"
BASE COUNT       85 a    117 c    113 g     60 t
ORIGIN

  Query Match             39.2%;  Score 196.4;  DB 14;  Length 375;
  Best Local Similarity   89.1%;  Pred. No. 1.3e-29;
  Matches  212;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy        235 GGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTCTTCGTGTGGCTCTGC 294
Db          2 GGCTTCGCCCTGCTGCGCCTGCTGGCCTGCCACCTGGGGCTCCTCTTCGTGTGGCTCTGC 61

Qy        295 CAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGGGCCGCGCCAGCACCT 354
Db         62 CAGCGCTTCTCCCGCGCCCTCATGGCCGCCAAGAGGAGCTCCGGGACCGCGCCGGCGCCC 121

Qy        355 GCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAGCGCGTCCGAGTCTTC 414
Db        122 GCCTCGCCCTCGACTCCAGCCCCCGGACCGGGTGGCGAGGCCGAGAGCGTCCGCGTCTTC 181

Qy        415 CACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
Db        182 CACAAGCAGGCCTTTGAGTACATCTCCATTGCCCTGCGCATCGACGAGGAAGAGAAAG 239


RESULT 14
AQ937698
LOCUS       AQ937698                 486 bp    DNA     linear   GSS 23-AUG-2000
DEFINITION  NB6-323R Human NotI clones Homo sapiens genomic, genomic survey
            sequence.
ACCESSION   AQ937698
VERSION     AQ937698.1  GI:7214076
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Zabarovsky,E.R., Gizatullin,R., Podowski,R.M., Zabarovska,V.V., Xie
            ,L., Muravenko,O.V., Kozyrev,S., Petrenko,L., Skobeleva,N., Li,J.,
            Protopopov,A., Kashuba,V., Ernberg,I., Winberg,G. and Wahlestedt,C.
  TITLE     NotI clones in the analysis of the human genome
  JOURNAL   Nucleic Acids Res. 28 (7), 1635-1639 (2000)
  MEDLINE   20175728
   PUBMED   10710430
COMMENT     Contact: Podowski RM
            Center for Genomics Research
            Karolinska Institute
            17177 Stockholm, Sweden
            Tel: +46-8-728-6372
            Fax: +46-8-337983
            Email: Raf.Podowski@cgr.ki.se
            Class: NotI site.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="Human NotI clones"
BASE COUNT       74 a    149 c    183 g     78 t      2 others
ORIGIN

  Query Match             37.6%;  Score 188.2;  DB 28;  Length 486;
  Best Local Similarity   90.7%;  Pred. No. 5.9e-28;
  Matches  244;  Conservative    0;  Mismatches   20;  Indels    5;  Gaps    4;

Qy          1 GGTTCCCGTCGGTCTGCGGGAGGC-GGGTTATGGCGGCGGCGGC-AGTGAGAGCTGTGAA 58
Db        220 GGTTCCCGTCGGTCTGCGGGAGGCAGGGTTATGGCGGCGGCGGCAAGTGAGAGCTGTGAA 279

Qy         59 TGAATTCTCCGGGTGGACGAGGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCGGTGC 118
Db        280 TGAATTCTCCGGGTGGACGATGGAAGAAGAAAGGCTCCGGCGGCGCCAGCAACCCTGTGC 339

Qy        119 CTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCCGCCGCCGGGCCGGCCCCTC 178
Db        340 CTCCCAGGCCTTCGNCCCCTTGCCTGGCCCCCGGCACTTCCGACGGCGGTCCGGTCCATC 399

Qy        179 CGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTTTGTAGGCT 238
```

```
Db        400 CGTCCGAGTC-CCGCATAAGCGGAACCTGTACTATTTCTCCTACCCGCTGTT--GTAGGC 456

Qy        239 TCGCGCTGCTGCGTTTGGTCGCCTTCCAC 267

Db        457 TCTCGCTGCTGCGTTTGGTCGCCTTNCAC 485


RESULT 15
AQ683149
LOCUS       AQ683149                 548 bp    DNA     linear   GSS 28-JUN-1999
DEFINITION  HS_5378_B2_C12_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=954 Col=24 Row=F, genomic survey sequence.
ACCESSION   AQ683149
VERSION     AQ683149.1  GI:5259055
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 548)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 954  row: F  column: 24
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 548.
FEATURES             Location/Qualifiers
     source          1..548
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=954 Col=24 Row=F"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
                     and partially digested with a combination of EcoRI and
                     EcoRI Methylase. Size selected DNA was cloned into the
                     pBACe3.6 vector at EcoRI sites"
BASE COUNT      128 a    159 c    142 g    115 t      4 others
ORIGIN

  Query Match             37.2%;  Score 186.6;  DB 28;  Length 548;
  Best Local Similarity   66.4%;  Pred. No. 1.2e-27;
  Matches  267;  Conservative    0;  Mismatches  135;  Indels    0;  Gaps    0;

Qy        100 GGCGCCAGCAACCCGGTGCCTCCCAGGCCTCCGCCCCCTTGCCTGGCCCCCGCCCCTCCC 159

Db         72 GGGATCATCGACCTGGGCACACTATTTCTCTCGTTCATCTTCCTGAGCCCACCCCCTCTA 131

Qy        160 GCCGCCGGGCCGGCCCCTCCGCCCGAGTCGCCGCATAAGCGGAACCTGTACTATTTCTCC 219

Db        132 GATGTTCGGCCAGTCAGTGTGNGCATAATGGGCCATAAGCATAACATAGGCTATTTCTAA 191

Qy        220 TACCCGCTGTTTGTAGGCTTCGCGCTGCTGCGTTTGGTCGCCTTCCACCTGGGGCTCCTC 279

Db        192 CACCCGGTGCATGTACGCTTCGAGCTGATGGGACCGAGTCCCTTCCAGCTGAGGCACCTC 251

Qy        280 TTCGTGTGGCTCTGCCAGCGCTTCTCCCGCGCCCTCATGGCAGCCAAGAGGAGCTCCGGG 339

Db        252 TTCGAGAGGCTCTGACATCGCTTCTCCCGAACCCTCATGGCAGGCAATAGGAGCTCCTGC 311

Qy        340 GCCGCGCCAGCACCTGCCTCGGCCTCGGCCCCGGCGCCGGTGCCGGGCGGCGAGGCCGAG 399

Db        312 GCCACACCATCACCTGGCTAAGCCTAGGGCCCAACACAAGTGCCGGGCGGTGATGCCGAC 371

Qy        400 CGCGTCCGAGTCTTCCACAAACAGGCCTTCGAGTACATCTCCATTGCCCTGCGCATCGAT 459

Db        372 CACGTTCCAGTCTTCACAAGACTAGCGTTACAGTTCATGTACATTGCTCTGCGCGTCCCA 431

Qy        460 GAGGATGAGAAAGGTAACTAGGGGGCTGGGGGAGGGGGCGGC 501

Db        432 GAGGATGAAAACAGCAACTAACCCGCTGGGGTACGGGGAAGC 473


Search completed: October 24, 2003, 22:51:04
Job time : 1049.9 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1381.32 Seconds
(without alignments)
14806.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: 1 taactgactctgctgtgttt.....aacatggtaaaaaccatgtcg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	500	100.0	101584	9	CNS01DS5	AL121655 BAC seque
2	500	100.0	110000	6	AX093471	AX093471 Sequence
3	500	100.0	110000	9	HSA246003	AJ246003 Homo sapi
4	500	100.0	155943	9	AC012364	AC012364 Homo sapi
c 5	203.8	40.8	133568	9	HS169I5	Z93015 Human DNA s
c 6	190.8	38.2	117751	9	AC020913	AC020913 Homo sapi
c 7	189.6	38.0	45195	9	AC090421	AC090421 Homo sapi
8	189.6	37.9	156589	9	HS537K23	AL034405 Human DNA
9	187.6	37.5	172837	9	AL450304	AL450304 Human DNA
10	181.4	36.3	169334	2	AC135837	AC135837 Papio anu
11	181.4	36.3	184341	2	AC135835	AC135835 Papio anu
12	179.4	35.9	89319	2	AL929255	AL929255 Homo sapi
c 13	179.4	35.9	97621	9	AL607089	AL607089 Human DNA
c 14	179.4	35.9	121362	9	AL359203	AL359203 Human DNA
15	179.4	35.9	145736	9	AL451139	AL451139 Human DNA
c 16	179.4	35.9	152428	9	AC114490	AC114490 Homo sapi
17	179.2	35.8	124337	9	HSBA436C9	AL121825 Human DNA
c 18	179	35.8	159130	9	AC091504	AC091504 Pan trogl
19	177.4	35.5	188546	2	AC022396	AC022396 Homo sapi
20	177.2	35.4	171309	9	AC012146	AC012146 Homo sapi
c 21	176.8	35.4	90214	9	AL356488	AL356488 Human DNA
22	176.8	35.4	190648	9	CNS01DXI	AL139317 Human chr
c 23	176.6	35.3	186902	9	AC010863	AC010863 Homo sapi
c 24	176.4	35.3	90143	9	AL669876	AL669876 Human DNA
c 25	176.4	35.3	151980	9	AC084209	AC084209 Homo sapi
26	176	35.2	106034	9	AC134028	AC134028 Homo sapi
c 27	176	35.2	154788	2	AC031984	AC031984 Homo sapi
c 28	176	35.2	186252	2	AC144518	AC144518 Pan trogl
c 29	175.8	35.2	110000	9	AE014305_0	AE014305 Homo sapi
c 30	175.8	35.2	164061	9	AL445188	AL445188 Human DNA
31	175.8	35.2	204439	9	AP000807	AP000807 Homo sapi
32	175.6	35.1	39729	6	AX647093	AX647093 Sequence
33	175.6	35.1	64475	9	AF283320S2	AF283321 Homo sapi
34	175.6	35.1	66933	6	AX277532	AX277532 Sequence
35	175.6	35.1	66933	6	AX418096	AX418096 Sequence
36	175.6	35.1	137206	2	AC024123	AC024123 Homo sapi
37	175.6	35.1	198720	9	AP002366	AP002366 Homo sapi
c 38	175.4	35.1	78167	9	AC092028	AC092028 Homo sapi
39	175	35.0	105087	9	AC093884	AC093884 Homo sapi
40	175	35.0	128152	9	AL136313	AL136313 Human DNA
41	174.8	35.0	59257	9	AC106019	AC106019 Homo sapi
42	174.8	35.0	110144	9	AC008053	AC008053 Homo sapi
c 43	174.8	35.0	130855	9	AC004089	AC004089 Homo sapi
c 44	174.8	35.0	188340	9	AC135178	AC135178 Homo sapi
45	174.8	35.0	189623	2	AC018640	AC018640 Homo sapi

ALIGNMENTS

RESULT 1
CNS01DS5
LOCUS CNS01DS5 101584 bp DNA linear PRI 16-APR-2002
DEFINITION BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 336P14 of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION AL121655
VERSION AL121655.1 GI:6002388
KEYWORDS SPG4 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101584)
AUTHORS Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D., Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,

```
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a novel AAA protein, is altered in the most frequent form
            of autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. (1999) In press
REFERENCE   2  (bases 1 to 101584)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .101584
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="336P14"
                     /clone_lib="CITB_978_SKB"
BASE COUNT    29860 a   22110 c   22047 g   27567 t
ORIGIN

  Query Match           100.0%;  Score 500;  DB 9;  Length 101584;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-115;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      59974 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60033

Qy         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60034 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 60093

Qy        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60094 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 60153

Qy        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60154 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 60213

Qy        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60214 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 60273

Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60274 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 60333

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60334 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 60393

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      60394 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 60453

Qy        481 AACATGGTAAAACCATGTCG 500
              ||||||||||||||||||||
Db      60454 AACATGGTAAAACCATGTCG 60473


RESULT 2
AX093471
LOCUS       AX093471               110000 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 1 from Patent WO0118198.
ACCESSION   AX093471
VERSION     AX093471.1  GI:13508911
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 1 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     exon            9332. .10471
     intron          10472. .33718
     exon            33719. .33805
     intron          33806. .35748
     exon            35749. .35832
     intron          35833. .45022
     exon            45023. .45118
     intron          45119. .60863
     exon            60864. .61051
     intron          61052. .61927
     exon            61928. .62061
     intron          62062. .62344
     exon            62345. .62438
     intron          62439. .73173
     exon            73174. .73248
     intron          73249. .74633
     exon            74634. .74705
     intron          74706. .82788
     exon            82789. .82864
     intron          82865. .83102
     exon            83103. .83194
     intron          83195. .83334
     exon            83335. .93414
     intron          83415. .88129
     exon            88130. .88172
     intron          88173. .89561
     exon            89562. .89641
     intron          89642. .91162
     exon            91163. .91233
     intron          91234. .93443
     exon            93444. .93484
     intron          93485. .100599
     exon            100600. .102009
BASE COUNT    30622 a   21640 c   22817 g   34921 t
ORIGIN

  Query Match           100.0%;  Score 500;  DB 6;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-115;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60

Qy         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120

Qy        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180

Qy        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240

Qy        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
```

```
Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360
              ||||  |||||| ||||| ||||| ||| ||||  || ||  || | |||  || |||
Db        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420
              ||||||  |||||| |||| ||| ||  ||  ||  |||  ||  ||
Db        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480
              |  |||||||| |||| ||||  |||| ||| |||| || ||| |||  ||
Db        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480

Qy        481 AACATGGTAAAACCATGTCG 500
              ||| |||||| ||||||
Db        481 AACATGGTAAAACCATGTCG 500

RESULT 3
HSA246003
LOCUS       HSA246003              110000 bp    DNA     linear   PRI 08-MAR-2000
DEFINITION  Homo sapiens Spast gene for spastin protein.
ACCESSION   AJ246003
VERSION     AJ246003.1  GI:6273492
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 110000)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The
            sequence is the result of the assembly of 2 BAC clones: R-336p14
            and
            563N4, respectively from RPCI-11 and CITB_978_SKB library.
FEATURES             Location/Qualifiers
     source          1..110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
                     /clone="R-336P14"
                     /clone_lib="RPCI-11"
                     /note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
                     , BAC 563N4 contains Spast gene: exon 4 to exon 17"
     gene            9932..102009
                     /gene="Spast"
     mRNA            9932..102009
                     /gene="Spast"
                     /product="spastin protein"
     exon            9932..10471
                     /gene="Spast"
                     /number=1
     5'UTR           9932..10056
                     /gene="Spast"
     CDS             join(10057..10471,33719..33805,35749..35832,45023..45118,
                     60864..61051,61928..62061,62345..62438,73174..73248,
                     74634..74705,82789..82864,83103..83194,83335..83414,
                     88130..88172,89562..89641,91163..91233,93444..93484,
                     100600..100722)
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60208.1"
                     /db_xref="GI:6273493"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     intron          10472..33718
                     /gene="Spast"
                     /number=1
     exon            33719..33805
                     /gene="Spast"
                     /number=2
     intron          33806..35748
                     /gene="Spast"
                     /number=2
     exon            35749..35832
                     /gene="Spast"
                     /number=3
     intron          35833..45022
                     /gene="Spast"
                     /number=3
     exon            45023..45118
                     /gene="Spast"
                     /number=4
     intron          45119..60863
                     /gene="Spast"
                     /number=4
     exon            60864..61051
                     /gene="Spast"
                     /number=5
     intron          61052..61927
                     /gene="Spast"
                     /number=5
     exon            61928..62061
                     /gene="Spast"
                     /number=6
     intron          62062..62344
                     /gene="Spast"
                     /number=6
     exon            62345..62438
                     /gene="Spast"
                     /number=7
     intron          62439..73173
                     /gene="Spast"
                     /number=7
     exon            73174..73248
                     /gene="Spast"
                     /number=8
     intron          73249..74633
                     /gene="Spast"
```

```
                     /number=8
     exon            74634. .74705
                     /gene="Spast"
                     /number=9
     intron          74706. .82788
                     /gene="Spast"
                     /number=9
     exon            82789. .82864
                     /gene="Spast"
                     /number=10
     intron          82865. .83102
                     /gene="Spast"
                     /number=10
     exon            83103. .83194
                     /gene="Spast"
                     /number=11
     intron          83195. .83334
                     /gene="Spast"
                     /number=11
     exon            83335. .83414
                     /gene="Spast"
                     /number=12
     intron          83415. .88129
                     /gene="Spast"
                     /number=12
     exon            88130. .88172
                     /gene="Spast"
                     /number=13
     intron          88173. .89561
                     /gene="Spast"
                     /number=13
     exon            89562. .89641
                     /gene="Spast"
                     /number=14
     intron          89642. .91162
                     /gene="Spast"
                     /number=14
     exon            91163. .91233
                     /gene="Spast"
                     /number=15
     intron          91234. .93443
                     /gene="Spast"
                     /number=15
     exon            93444. .93484
                     /gene="Spast"
                     /number=16
     intron          93485. .100599
                     /gene="Spast"
                     /number=16
     exon            100600. .102009
                     /gene="Spast"
                     /number=17
     3'UTR           100721. .102009
                     /gene="Spast"
     polyA_signal    101973. .101978
                     /gene="Spast"
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match           100.0%;  Score 500;  DB 9;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-115;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60

Qy         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120

Qy        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180

Qy        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240

Qy        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300

Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480

Qy        481 AACATGGTAAAACCATGTCG 500
              ||||||||||||||||||||
Db        481 AACATGGTAAAACCATGTCG 500


RESULT 4
AC012364
LOCUS       AC012364              155943 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-444D15 from 2, complete sequence.
ACCESSION   AC012364
VERSION     AC012364.7  GI:14589736
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155943)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 155943)
  AUTHORS   Ali,J. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11 444D15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 155943)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:11602953.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0444D15
```

```
            ---------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Csoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1999) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 59:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-431P13; the clone sequenced
            to the right is RP11-78E13.  Actual start of this clone is at base
            position 1 of RP11-444D15; actual end is at base position 155943 of
            RP11-444D15.

            Data from AC011232 and AC010981 was used to finish this clone,
            AC012364.  The sequence fidelity between bases 112626 to 112643 can
            not be guaranteed due to an unresolved homopolymeric run.  The
            sequence between 112648 to 112663 is single stranded.
FEATURES             Location/Qualifiers
     source          1..155943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-444D15"
                     /clone_lib="RPCI-11"
     repeat_region   310..617
                     /rpt_family="Alu"
     repeat_region   593..617
                     /rpt_family="AT_rich"
     repeat_region   728..748
                     /rpt_family="AT_rich"
     repeat_region   732..1026
                     /rpt_family="Alu"
     repeat_region   1169..1240
                     /rpt_family="(TATATG)n"
     repeat_region   1610..2503
                     /rpt_family="L2"
     repeat_region   2504..2776
                     /rpt_family="Alu"
     repeat_region   2777..3498
                     /rpt_family="L2"
     repeat_region   3590..3742
                     /rpt_family="MIR"
     repeat_region   4452..4765
                     /rpt_family="Alu"
     misc_feature    4550..4975
                     /note="match to EST AA632695 (NID:g2556109) np83b02.s1"
     misc_feature    4661..4692
                     /note="match to EST BG571163 (NID:g13578829)"
     repeat_region   4970..5239
                     /rpt_family="Alu"
     repeat_region   5211..5238
                     /rpt_family="(CAAAAA)n"
     misc_feature    5234..5328
                     /note="match to EST BG571163 (NID:g13578829)"
     misc_feature    5242..5328
                     /note="match to EST BG757662 (NID:g14068315)"
     misc_feature    5247..5328
                     /note="match to EST AU121379 (NID:g10936614)"
     misc_feature    5247..5328
                     /note="match to EST BF205430 (NID:g11099016)"
     misc_feature    5247..5328
                     /note="match to EST BG033287 (NID:g12425431)"
     misc_feature    5247..5328
                     /note="similar to Mus musculus EST AA619874 (NID:g2523753)
                     vl58g02.s1"
     misc_feature    5247..5328
                     /note="similar to Mus musculus EST AW120623
                     (NID:g6095956)"
     repeat_region   5740..6346
                     /rpt_family="L2"
     repeat_region   6347..6438
                     /rpt_family="AcHobo"
     repeat_region   6453..6777
                     /rpt_family="L2"
     repeat_region   6938..7406
                     /rpt_family="L1"
     repeat_region   7415..8075
                     /rpt_family="ERV1"
     misc_feature    7900..8421
                     /note="match to EST AW994939 (NID:g8255173)"
     repeat_region   8087..8245
                     /rpt_family="L1"
     repeat_region   8447..8524
                     /rpt_family="MER2_type"
     repeat_region   8525..8799
                     /rpt_family="Alu"
     repeat_region   8800..9067
                     /rpt_family="MER2_type"
     repeat_region   9917..10069
                     /rpt_family="MER1_type"
     misc_feature    10293..10786
                     /note="match to EST N54249 (NID:g1195415) yv65c12.s1"
     repeat_region   10566..10753
                     /rpt_family="Alu"
     repeat_region   10820..11873
                     /rpt_family="L1"
     repeat_region   11588..11607
                     /rpt_family="(A)n"
     repeat_region   11863..11888
                     /rpt_family="AT_rich"
     repeat_region   11887..12044
                     /rpt_family="Tc2"
     repeat_region   12045..12345
                     /rpt_family="Alu"
     repeat_region   12382..12694
                     /rpt_family="Alu"
     repeat_region   12666..12727
                     /rpt_family="A-rich"
     repeat_region   12743..13296
                     /rpt_family="Tc2"
     repeat_region   13463..13492
                     /rpt_family="(T)n"
     misc_feature    13463..13478
                     /note="similar to Homo sapiens EST BG110835
                     (NID:g12604341)"
     misc_feature    13820..14405
                     /note="match to EST BG434159 (NID:g13340653)"
     repeat_region   14067..14097
                     /rpt_family="AT_rich"
     repeat_region   14793..15032
```

```
                     /rpt_family="L1"
     repeat_region   15033..15330
                     /rpt_family="Alu"
     repeat_region   15331..15399
                     /rpt_family="L1"
     repeat_region   15400..15700
                     /rpt_family="Alu"
     misc_feature    15400..15410
                     /note="match to EST BG576844 (NID:g13584497)"
     repeat_region   15701..15745

  Query Match             100.0%;  Score 500;  DB 9;  Length 135943;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-115;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60
Db     115737 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 115796

Qy         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120
Db     115797 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 115856

Qy        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTGGTTTTGTTT 180
Db     115857 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTGGTTTTGTTT 115916

Qy        181 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240
Db     115917 TTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 115976

Qy        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
Db     115977 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 116036

Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360
Db     116037 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 116096

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420
Db     116097 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 116156

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480
Db     116157 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 116216

Qy        481 AACATGGTAAAACCATGTCG 500
Db     116217 AACATGGTAAAACCATGTCG 116236


RESULT 5
HS169I5/c
LOCUS       HS169I5          133568 bp    DNA     linear   PRI 13-NOV-2000
DEFINITION  Human DNA sequence from clone RP1-169I5 on chromosome Xp11.
            Contains ESTs, STSs, GSSs and CpG islands. Contains the 3' part of
            the DDX3 gene for DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3
            and a novel gene for a leucine rich protein, complete sequence.
ACCESSION   Z93015
VERSION     Z93015.9  GI:9929207
KEYWORDS    HTG; CpG island; DDX3; DEAD box; leucine rich protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 133568)
  AUTHORS   Pavitt,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9798426.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP1-169I5 is from the library RPCI-1 constructed at the Roswell
            Park Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP1-169I5.
FEATURES             Location/Qualifiers
     source          1..133568
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="p11"
                     /clone="RP1-169I5"
                     /clone_lib="RPCI-1"
     repeat_region   1..78
                     /note="AluSp/q repeat: matches 234..311 of consensus"
     gene            541..789
                     /gene="DDX3"
     mRNA            <541..789
                     /gene="DDX3"
                     /product="dJ169I5.1 (DEAD/H (Asp-Glu-Ala-Asp/His) box
                     polypeptide 3)"
                     /note="match: cDNAs: Em:AF000983"
                     /evidence=not_experimental
     misc_feature    638..736
                     /gene="DDX3"
                     /note="match: STS: Em:G34988"
     polyA_signal    713..718
                     /gene="DDX3"
     polyA_site      789
                     /gene="DDX3"
     repeat_region   807..884
                     /note="AluJ/FLAM repeat: matches 2..79 of consensus"
     repeat_region   899..1025
                     /note="L1MC5 repeat: matches 7741..7890 of consensus"
     repeat_region   3392..3431
                     /note="L1ME repeat: matches 5776..5817 of consensus"
     repeat_region   3743..3818
                     /note="L1ME repeat: matches 5709..5776 of consensus"
     repeat_region   3838..3918
                     /note="L1MC5 repeat: matches 7833..7916 of consensus"
     repeat_region   3944..4108
                     /note="MER5B repeat: matches 1..161 of consensus"
     repeat_region   5253..5783
                     /note="L1ME1 repeat: matches 5669..6163 of consensus"
     repeat_region   6086..6477
                     /note="L1ME1 repeat: matches 5182..5669 of consensus"
     repeat_region   6778..6881
                     /note="L1ME1 repeat: matches 5086..5182 of consensus"
     repeat_region   7186..7530
                     /note="L1 repeat: matches 3143..3507 of consensus"
     repeat_region   7854..8044
```

```
                     /note="AluSg/x repeat: matches 118..308 of consensus"
     misc_feature    complement(11172..11529)
                     /note="match: GSS: Em:B93158"
     misc_feature    complement(11793..12084)
                     /note="match: GSS: Em:AQ201865"
     misc_feature    complement(11813..12083)
                     /note="match: STS: Em:G21430 Em:G33608"
     misc_feature    complement(11822..12084)
                     /note="match: GSS: Em:AQ198952
                     match: STS: Em:G52853"
     repeat_region   14958..15328
                     /note="L1ME3 repeat: matches 5589..5962 of consensus"
     repeat_region   15766..15958
                     /note="L1ME3 repeat: matches 5972..6157 of consensus"
     misc_feature    16159..16317
                     /note="match: GSS: Em:B72913"
     repeat_region   19667..19698
                     /note="16 copies 2 mer aa 93% conserved"
     repeat_region   20404..20431
                     /note="14 copies 2 mer tt 96% conserved"
     repeat_region   20899..21035
                     /note="SVA repeat: matches 1..133 of consensus"
     repeat_region   21444..21589
                     /note="MER4C repeat: matches 302..448 of consensus"
     repeat_region   22551..22862
                     /note="MER4C repeat: matches 2..318 of consensus"
     repeat_region   23001..23176
                     /note="LTR23 repeat: matches 59..247 of consensus"
     repeat_region   23506..23647
                     /note="LTR23 repeat: matches 247..437 of consensus"
     repeat_region   23838..24020
                     /note="AluSg/x repeat: matches 130..310 of consensus"
     repeat_region   24292..24564
                     /note="LTR39 repeat: matches 541..792 of consensus"
     repeat_region   24878..25143
                     /note="LTR39 repeat: matches 212..541 of consensus"
     repeat_region   26355..26499
                     /note="MER41A repeat: matches 1..152 of consensus"
     repeat_region   26810..27173
                     /note="MER41A repeat: matches 152..478 of consensus"
     repeat_region   27471..27544
                     /note="MER41A repeat: matches 478..554 of consensus"
     misc_feature    29053..29217
                     /note="match: GSS: Em:AC029328"
     repeat_region   30341..30572
                     /note="LTR16C repeat: matches 157..387 of consensus"
     repeat_region   34429..34515
                     /note="MLT2E repeat: matches 44..131 of consensus"
     misc_feature    34772..35018
                     /note="match: GSS: Em:AQ469146"
     misc_feature    complement(34864..35259)
                     /note="match: GSS: Em:AQ432633"
     misc_feature    35276..35677
                     /note="match: GSS: Em:AQ150455"
     repeat_region   35690..35711
                     /note="11 copies 2 mer ct 100% conserved"
     repeat_region   35725..35898
                     /note="AluSg/x repeat: matches 127..301 of consensus"
     repeat_region   37022..37073
                     /note="26 copies 2 mer ta 73% conserved"
     repeat_region   38678..39824
                     /note="L1PA10 repeat: matches 5003..6159 of consensus"
     repeat_region   45727..45766
                     /note="20 copies 2 mer ac 97% conserved"
     repeat_region   46967..47132
                     /note="LTR8 repeat: matches 543..691 of consensus"
     repeat_region   47425..47563
                     /note="LTR8 repeat: matches 379..543 of consensus"
     repeat_region   47581..47728
                     /note="LTR8 repeat: matches 1..148 of consensus"
     repeat_region   50494..50815
                     /note="MER1B repeat: matches 1..337 of consensus"
     repeat_region   53887..54081
                     /note="L1ME3A repeat: matches 5971..6162 of consensus"
     repeat_region   54359..54543
                     /note="L1ME3A repeat: matches 5777..5971 of consensus"
     repeat_region   54547..54693
                     /note="AluJo/FRAM repeat: matches 146..292 of consensus"
     repeat_region   54704..54792
                     /note="L1ME3A repeat: matches 5699..5787 of consensus"
     repeat_region   56108..56133
                     /note="L1M1 repeat: matches 5255..5280 of consensus"
     repeat_region   56441..56596
                     /note="L1M1 repeat: matches 5280..5433 of consensus"
     repeat_region   56912..56970
                     /note="L1ME3A repeat: matches 5741..5800 of consensus"
     repeat_region   57042..57298
                     /note="L1ME3A repeat: matches 5785..6039 of consensus'
     repeat_region   57991..58347
                     /note="Charlie1b repeat: matches 1..372 of consensus"
     repeat_region   58664..58703
                     /note="Charlie1b repeat: matches 372..412 of consensus"
     repeat_region   59021..59167
                     /note="Charlie1b repeat: matches 412..523 of consensus"
     repeat_region   59887..60025
                     /note="TIGGER2 repeat: matches 2391..2529 of consensus"
     repeat_region   60487..60710
                     /note="L1MC/D repeat: matches 5551..5800 of consensus"
     repeat_region   61616..61690
                     /note="L1MC/D repeat: matches 5805..5823 of consensus"
     repeat_region   62967..64943
                     /note="L1PA3 repeat: matches 4170..6146 of consensus"
     repeat_region   64934..65145
                     /note="L1MC5 repeat: matches 7673..7917 of consensus"
     repeat_region   67893..68097
                     /note="Charlie1b repeat: matches 323..522 of consensus"
     repeat_region   68415..68684
                     /note="Charlie1b repeat: matches 78..323 of consensus"

  Query Match             40.8%;  Score 203.8;  DB 9;  Length 133568;
  Best Local Similarity   75.5%;  Pred. No. 6.2e-41;
  Matches  253;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        165 AGTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTCTAGATGGAGTC 224
Db       8065 ATTTTTATTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTCTGAGACAGAGTC 8006

Qy        225 TTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACC 284
Db       8005 TCGCTCTGTCACCCAGGTTGGAGTGCAGTGGCACAATCTCAGCTCACTACAACCTCCGCC 7946

Qy        285 TTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGA 344
Db       7945 TCCTGGGTTTAAGCAATTCTCCTGCCTCAGCCTCCCAAGTTGCTGGGACTACAGGTGCAC 7886

Qy        345 GCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCA 404
Db       7885 GCCACCACGCCTGGCTAATTTTTGTATTGTTATGAGATACAGGCCAGGCGCTGTGGCTCA 7826

Qy        405 TATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTT 464
Db       7825 CACCTGTAATCTCAACACTTTGGGAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTT 7766

Qy        465 CAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
Db       7765 CTAGACCAGTCTGGCCAACATGGTGAAACCCTGTC 7731


RESULT 6
AC020913/c
LOCUS       AC020913              117751 bp    DNA     linear   PRI 14-JUL-2002
DEFINITION  Homo sapiens chromosome 19 clone CTD-3032J10, complete sequence.
ACCESSION   AC020913
VERSION     AC020913.9  GI:21747450
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   5  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   6  (bases 1 to 117751)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 14, 2002 this sequence version replaced gi:21204238.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.7% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES             Location/Qualifiers
     source          1..117751
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTD-3032J10"
BASE COUNT    28597 a  31370 c  30506 g  27278 t
ORIGIN

  Query Match             38.2%;  Score 190.8;  DB 9;  Length 117751;
  Best Local Similarity   72.8%;  Pred. No. 1.2e-37;
  Matches  246;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy        162 GGGAGTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGA 221
              ||   |||   |  | ||   ||  | |||  |||||||||| |||  |||  |    |
Db      76941 GGATATTTATTTAGTTTTAAATTTATTTTTATTTTTATTTTATTTTTTTTTGGAAACAGT 76882

Qy        222 GTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCC 281
              |||||||||||||||||||||||| ||| ||  | |||| |||| |||||||||||||
Db      76881 GTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGGTGCGATCTCAGCTCACTGCAACCTCC 76822

Qy        282 ACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAA 341
              || || | |||||||| || ||||||||| | |  |||||||| ||||  |
Db      76821 ACTTTCCAGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGACA 76762

Qy        342 TGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGC 401
              ||     ||    . |||  |  ||| .        ||||  | || ||  |
Db      76761 TGCATCACCCTGCTCAGCCTATTTATTTTTTAAATTCCATGACAGGCTGGGCGCGGTGGC 76702

Qy        402 CCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCGATTACTTGAGGTCAGGA 461
              |||  |||| ||||||||||||  ||| |  ||   ||| ||  || |||  : |
Db      76701 TCATGCCTGTAATCCCAGCACTTTCGGAGGCTGAGCAGGGCAGATCACCTGAGGTAGGGA 76642

Qy        462 GTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
              ||||||||||||  ||| |||||| ||||  |||
Db      76641 GTTCAAGACCAGCCTGGCCAACATGGTGAAACCCCGTC 76604


RESULT 7
AC090421/c
LOCUS       AC090421               45195 bp    DNA     linear   PRI 29-APR-2003
DEFINITION  Homo sapiens BAC clone RP11-441N19 from 7, complete sequence.
ACCESSION   AC090421
VERSION     AC090421.5  GI:15638955
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 45195)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 45195)
  AUTHORS   Dixon,R., Du,F., Elliott,G. and Haakenson,W.
  TITLE     The sequence of Homo sapiens BAC clone RP11-441N19
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 45195)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 45195)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 45195)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   6  (bases 1 to 45195)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 18, 2001 this sequence version replaced gi:15148341.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0441N19
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            The sequence of this clone was established as part of a mapping and
```

```
            sequencing collaboration between the NHGRI Chromosome 7 Mapping
            Project (Eric D. Green, Director), John D. McPherson in the
            Department of Genetics (Washington University), and the Washington
            University Genome Sequencing Center. For additional information
            about the map position of this sequence, see
            http://www.nhgri.nih.gov/DIR/GTB/CHR7 , send
            mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries. Genomics 51:1-8. The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at http://www.chori.org
            VECTOR:  pBACe3.6

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP4-562A11, 2000 bp overlap the
            clone sequenced to the right is RP5-1087M19, 2000 bp overlap.
            Actual start of this clone is at base position 40901 of RP4-562A11
            actual end is at base position 13013 of RP5-1087M19.
FEATURES             Location/Qualifiers
     source          1..45195
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-441N19"
                     /clone_lib="RPCI-11"
     repeat_region   143..279
                     /rpt_family="Alu"
     repeat_region   280..577
                     /rpt_family="Alu"
     repeat_region   578..882
                     /rpt_family="Alu"
     repeat_region   883..975
                     /rpt_family="Alu"
     repeat_region   987..1108
                     /rpt_family="GA-rich"
     repeat_region   1302..1600
                     /rpt_family="Alu"
     repeat_region   2567..2867
                     /rpt_family="Alu"
     repeat_region   3096..3425
                     /rpt_family="Alu"
     repeat_region   4031..4252
                     /rpt_family="MIR"
     repeat_region   4377..4675
                     /rpt_family="Alu"
     repeat_region   4680..4707
                     /rpt_family="(CA)n"
     repeat_region   4708..4759
                     /rpt_family="(TATG)n"
     repeat_region   4761..5051
                     /rpt_family="Alu"
     repeat_region   5084..5384
                     /rpt_family="Alu"
     repeat_region   5394..5562
                     /rpt_family="Alu"
     repeat_region   5773..6016
                     /rpt_family="MaLR"
     repeat_region   6114..6253
                     /rpt_family="Alu"
     repeat_region   6728..7034
                     /rpt_family="Alu"
     repeat_region   7136..7266
                     /rpt_family="MIR"
     repeat_region   7925..8187
                     /rpt_family="Alu"
     repeat_region   8424..8697
                     /rpt_family="Alu"
     repeat_region   9038..9170
                     /rpt_family="Alu"
     repeat_region   9171..9449
                     /rpt_family="Alu"
     repeat_region   9450..9627
                     /rpt_family="Alu"
     repeat_region   9755..9969
                     /rpt_family="L1"
     variation       9877
                     /allele="A"
                     /allele="G"
                     /db_xref="dbSNP:1130039"
     repeat_region   10365..10493
                     /rpt_family="L2"
     repeat_region   10719..10745
                     /rpt_family="(CAAAAA)n"
     repeat_region   10747..11042
                     /rpt_family="Alu"
     repeat_region   11424..11717
                     /rpt_family="Alu"
     repeat_region   11718..12020
                     /rpt_family="Alu"
     repeat_region   12032..12338
                     /rpt_family="Alu"
     repeat_region   12457..12741
                     /rpt_family="Alu"
     repeat_region   12743..13056
                     /rpt_family="Alu"
     repeat_region   13057..13232
                     /rpt_family="Alu"
     repeat_region   13234..13431
                     /rpt_family="Alu"
     repeat_region   13467..13765
                     /rpt_family="Alu"
     repeat_region   14164..14450
                     /rpt_family="Alu"
     repeat_region   14750..15010
                     /rpt_family="Alu"
     repeat_region   15011..15263
                     /rpt_family="GA-rich"
     repeat_region   15550..15895
                     /rpt_family="L1"
     repeat_region   15905..16088
                     /rpt_family="Alu"
     repeat_region   16110..16228
                     /rpt_family="Alu"
     repeat_region   16229..16525
                     /rpt_family="Alu"
     repeat_region   16526..16727
                     /rpt_family="Alu"
     repeat_region   16755..17036
                     /rpt_family="L1"
     repeat_region   17048..17245
                     /rpt_family="MIR"
     repeat_region   17562..17664
                     /rpt_family="MaLR"
     repeat_region   17697..17827
                     /rpt_family="Alu"
     repeat_region   17857..17988
                     /rpt_family="MaLR"

  Query Match             38.0%;  Score 189.8;  DB 9;  Length 45195;
  Best Local Similarity   75.5%;  Pred. No. 2e-37;
  Matches  250;  Conservative    0;  Mismatches   77;  Indels    4;  Gaps    1;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232
               |||   |||  ||||: |||   |||  .||    |   |  :.   ||    :||||   |||||
Db      41368 TTTGTGTTTTTTGTTTTTTGTTTTGTTTTGTTTTGTTTTTTGAGATGGAGTCTCACTCTG 41309

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
                 ||:||||   |||  |.||  :|      ||||:|||  |  |||||:.  |  ||.   .||||
Db      41308 TCGCCCAGGCTGGAGTGCAGTGGTGCAGTCTCAGCTCACTGCAACCTCCGCCTCCCGGGT 41249
```

```
Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
              ||||||||   ||||||||  |  |  |||||  |||  |||   |||||  |    |,  ||
Db      41248 TCAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCCGCCACCAT 41189

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTTAAA----TTTTGAGGCCAGGTGCAGTGGCCCATATC 408
                || |||    |  |   |       ||:      ,.||  .||   |||||||    |      .
Db      41188 GCCCGGCTAATTTTTGTATTCAATAAATTGTTTTTCAGGCTAGGTGCAGTGGCTCACTCC 41129

Qy        409 TGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAG 468
              ||| ||||||||||||||||||||  |  |||  |||  ||  |  |  |||||||||||||    ||
Db      41128 TGTAATCCCAGCACTTTGGGAGGCTGAGGTGGCAGACTGCCTGAGGTCAGGAGTTCGAG 41069

Qy        469 ACCAGCCAGGCCAACATGGTAAAACCATGTC 499
              ||||.|| |:| |||||  ||  :|||    ||
Db      41068 ACCAGCCTGGCTAACATAGTGAAACCCCATC 41038
```

```
RESULT 8
HS537K23
LOCUS       HS537K23               156589 bp    DNA     linear   PRI 02-JAN-2002
DEFINITION  Human DNA sequence from clone RP4-537K23 on chromosome Xq25-26.1
            Contains the 3' end of the gene for CGI-89 protein, a pseudogene
            similar to SALL1 (SAL(drosophila)-like 1), the 5' end of a gene for
            KIAA0266 protein and CpG islands, complete sequence.
ACCESSION   AL034405
VERSION     AL034405.16  GI:9798442
KEYWORDS    HTG; CpG island; KIAA0266.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156589)
  AUTHORS   Grafham,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8230880.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            This sequence is the entire insert of clone RP4-537K23 The true
            left end of clone RP3-426N21 is at 105462 in this sequence. This
            sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. RP4-537K23 is from
            the library RPCI-4 constructed by the group of Pieter de Jong. For
            further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2.
FEATURES             Location/Qualifiers
     source          1..156589
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q25-26.1"
                     /clone="RP4-537K23"
                     /clone_lib="RPCI-4"
     misc_feature    52..528
                     /note="match: GSS: Em:AQ426659"
     repeat_region   1863..2169
                     /note="L1MA8 repeat: matches 5978..6288 of consensus"
     repeat_region   9705..9752
                     /note="12 copies 4 mer caca 77% conserved"
     repeat_region   9706..9745
                     /note="20 copies 2 mer ac 82% conserved"
     misc_feature    complement(11187..11626)
                     /note="match: GSS: Em:AQ175289"
     misc_feature    11627..12178
                     /note="match: GSS: Em:AQ322765"
     gene            complement(11770..11993)
                     /gene="dJ537K23.1"
     CDS             complement(<11770..>11993)
                     /gene="dJ537K23.1"
                     /note="match: proteins: Tr:Q9Y397"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ537K23.1 (CGI-89 protein)"
                     /protein_id="CAD20046.1"
                     /db_xref="GI:18056468"
                     /translation="PLFLLSCNRKTHFGAGIPIMSVMVVRKKVTRKWEKLPGRNTFCC
                     DGRVMMARQKGIFYLTLFLILGTCTLFFAF"
     misc_feature    11984..12201
                     /note="match: GSS: Em:AQ551453"
     misc_feature    13341..14223
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   13487..13534
                     /note="12 copies 4 mer caca 75% conserved"
     repeat_region   14190..14237
                     /note="24 copies 2 mer ca 77% conserved"
     misc_feature    complement(15133..15649)
                     /note="match: GSS: Em:AQ056705"
     repeat_region   15199..15296
                     /note="2 copies 49 mer 90% conserved"
     misc_feature    complement(16388..16880)
                     /note="match: GSS: Em:B42283"
     repeat_region   16603..16636
                     /note="17 copies 2 mer ga 82% conserved"
     repeat_region   16608..16635
                     /note="7 copies 4 mer agaa 92% conserved"
     repeat_region   16784..16839
                     /note="28 copies 2 mer tg 85% conserved"
     repeat_region   19661..19734
                     /note="37 copies 2 mer cc 67% conserved"
     misc_feature    19897..21404
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   21979..22026
                     /note="24 copies 2 mer aa 81% conserved"
     misc_feature    22693..22917
                     /note="match: GSS: Em:AQ734022"
     misc_feature    22923..23204
                     /note="match: GSS: Em:AQ734022"
     repeat_region   23648..29352
                     /note="L1PA5 repeat: matches 433..6145 of consensus"
     misc_feature    complement(29592..29817)
                     /note="match: STS: Em:G15426"
     misc_feature    29621..29868
                     /note="match: GSS: Em:AF004855"
     misc_feature    complement(29629..29769)
                     /note="match: STS: Em:G20770"
     repeat_region   30275..30304
                     /note="10 copies 3 mer caa 93% conserved"
     misc_feature    31160..31526
                     /note="match: STS: Em:G13992"
     repeat_region   33448..33469
                     /note="11 copies 2 mer aa 100% conserved"
```

```
repeat_region   36173..36202
                /note="15 copies 2 mer ac 96% conserved"
repeat_region   36175..36202
                /note="7 copies 4 mer acac 100% conserved"
repeat_region   36331..36366
                /note="18 copies 2 mer tt 93% conserved"
repeat_region   36381..36411
                /note="L1PA7 repeat: matches 6095..6125 of consensus"
repeat_region   36411..36634
                /note="L1PA7 repeat: matches 5731..5954 of consensus"
gene            41301..47561
                /gene="dJ537K23.2"
                /pseudo
CDS             41301..47561
                /gene="dJ537K23.2"
                /note="dJ537K23.2 (similar to SALL1 (SAL(drosophila)-like
                1))
                match: proteins: Tr:Q91929 Tr:Q9PVN6 Tr:Q62255 Tr:C12958
                Tr:Q9W7G2 Tr:Q9PVN5 Tr:O94892 Tr:Q9QX96 Tr:P91630
                Sw:P39806 Tr:Q9Y2H8 Sw:P18746 Sw:P15622 Tr:Q9VKB3
                Sw:P39770 Tr:Q62510 Sw:P17032 Sw:P18730 Tr:Q9Y6Q3
                Sw:Q16600 Tr:Q17396 Sw:P10077 Tr:O14961 Tr:Q9VG72
                Tr:O61362 Tr:O75123 Sw:P10075 Tr:Q9Y4G1 Sw:P17141
                Tr:Q9UJN7 Tr:Q9Y5A5 Sw:P51523 Tr:O60304 Sw:P10754
                Tr:Q62517 Tr:O00146 Sw:P18713 Sw:Q62981 Tr:Q9UJW8
                Sw:P52736 Sw:P18753 Tr:Q9Y2P7 Sw:Q62513 Tr:O90658
                Sw:P52738 Tr:Q9W3S1 Tr:Q94178 Sw:P24399 Wp:C55B7 Tr:Q9VDQ6
                Sw:O95125 Tr:Q9UF97 Sw:P18724 Sw:P15619 Tr:O57415
                Tr:Q9WVL8 Sw:P18729 Tr:Q62886 Tr:Q9Y3M9 Tr:Q9Y467
                Wp:Y55F3AM Tr:Q9R161 Sw:P18747 Tr:Q9UIW8 Sw:P18749
                Tr:O74256 Sw:O14628 Sw:Q02386 Tr:O95878 Sw:P18736
                Sw:P18738"
                /pseudo
                /codon_start=1
                /evidence=not_experimental
misc_feature    41524..41878
                /gene="dJ537K23.2"
                /note="match: GSS: Em:AQ341439"
                /pseudo
repeat_region   42823..43053
                /note="MER41-internal repeat: matches 2967..3198 of
                consensus"
repeat_region   43352..43429
                /note="MER41-internal repeat: matches 3198..3275 of
                consensus"
repeat_region   43683..44388
                /note="MER41-internal repeat: matches 3275..3941 of
                consensus"
repeat_region   46241..46953
                /note="LTR1 repeat: matches 1..785 of consensus"
repeat_region   47857..48721
                /note="L1PB3 repeat: matches 5293..6300 of consensus"
repeat_region   48839..48890
                /note="26 copies 2 mer aa 73% conserved"
repeat_region   52743..52886
                /note="L1MA8 repeat: matches 6142..6285 of consensus"
repeat_region   57849..57901
                /note="AluJ/monomer repeat: matches 250..302 of consensus"
repeat_region   58354..58741
                /note="THE1C repeat: matches 1..371 of consensus"
repeat_region   63307..63348
                /note="21 copies 2 mer tt 78% conserved"
misc_feature    63539..63588
                /note="Single clone region"
repeat_region   65820..65929
                /note="AluJ/F(R)AM repeat: matches 149..297 of consensus"
repeat_region   66814..67015
                /note="MER63A repeat: matches 1..210 of consensus"
repeat_region   71827..71896
                /note="2 copies 35 mer 90% conserved"
misc_feature    75291..75714
                /note="match: GSS: Em:AQ172817"
misc_feature    75766..76167
```

```
                /note="CpG island"

  Query Match             37.9%;  Score 189.6;  DB 9;  Length 156589;
  Best Local Similarity   75.6%;  Pred. No. 2.3e-37;
  Matches  248;  Conservative    0;  Mismatches   79;  Indels    1;  Gaps    1;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232
              |  |  : |||:'| |  ': ||  |||   ||  |.|.  ||' ' .||:|'  .,|||
Db     113262 TATTTGTGTTTTGTTTTGTTTTTGTTTTTGTTTTTGTTTTTGAGACGGAGTCTCACTCTG 113321

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
              |  |||:|'||:.  ||':|||'   ||| :'  ||| : ;||||':   ||'  | | .
Db     113322 TCACCCAGGCTGGAGTGCAATGGCGTCATCTCACCTCACTGCAACCTCTGCCTCCCAGAT 113381

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
              |':||||'': |':|'  |'|  ;' || |  ||||:': | ||| !  |||'    ||
Db     113382 TCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGAGACACCAC 113441

Qy        353 ACCTGGCCTGGT-TTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGT 411
              |    |:'|  :     |.' |||  , |         || : | ||    ||||     ||.
Db     113442 ATCCAGCCTTTTAAAATTTATTTTAATAATACTTGGCCGGATGTGGTGGCTCATGCCTGT 113501

Qy        412 GATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACC 471
               . |||  |||:||||;  !  ;:|| | | |||  '   |||:', !|||; .  ||'||
Db     113502 AATCCCAACACTTTGGGAGGCTGAGGCTGGCAGATCATCTGAGGTCAGGAGTTTGAGACC 113561

Qy        472 AGCCAGGCCAACATGGTAAAACCATGTC 499
              |  | |.:|||:||||  ||||'|||
Db     113562 AGCCTGGCCAACATGGTGAAACCCTGTC 113589


RESULT 9
AL450304
LOCUS       AL450304           172837 bp    DNA     linear   PRI 20-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-505O17 on chromosome 1, complete
            sequence.
ACCESSION   AL450304
VERSION     AL450304.33  GI:17976576
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172837)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 21, 2001 this sequence version replaced gi:17384096.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-505O17 is from the library RPCI-11.2 constructed by the group
```

```
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-505O17 The true
            left end of clone RP11-181K3 is at 147571 in this sequence. The
            true right end of clone RP11-333G16 is at 119781 in this sequence.
FEATURES             Location/Qualifiers
     source          1..172837
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-505O17"
                     /clone_lib="RPCI-11.2"
     misc_feature    18441..19559
                     /note="Sequence from overlapping clone RP11-333G16
                     (AL590620). Assembly confirmed by restriction digest."
     misc_feature    20613..20657
                     /note="Sequence from overlapping clone RP11-333G16
                     (AL590620). Assembly confirmed by restriction digest."
     misc_feature    26437..26486
                     /note="Sequence from overlapping clone RP11-333G16
                     (AL590620). Assembly confirmed by restriction digest."
     misc_feature    99691..100086
                     /note="Single clone region. Assembly confirmed by
                     restriction digest data."
     misc_feature    99946..100002
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
     misc_feature    100203..100919
                     /note="Sequence from overlapping clone RP11-333G16
                     (AL590620). Assembly confirmed by restriction digest."
     misc_feature    127409..127498
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
BASE COUNT    49503 a  33911 c  34461 g  54962 t
ORIGIN

  Query Match             37.5%;  Score 187.6;  DB 9;  Length 172837;
  Best Local Similarity   75.2%;  Pred. No. 7.5e-37;
  Matches  248;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    1;

Qy        170 TGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCT 229

Db      43942 TGAGTTTCCCATTTTGGTGTTTTTGTTGTTTTGTTCTTTTGTTTAAGAAAGAGTCTCACT 44001

Qy        230 CTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCG 289

Db      44002 CTGTTGTCCAGGCTGGAGGGCAGTGGTGCAGTCTCAGCTCACTGCAACCTCCGCCTTCTG 44061

Qy        290 GGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGC 349

Db      44062 GGTTCAAGCAATTCTCATGCCTCAGACTCCCACGTAGTTGGGATTACAGGCATGAGACAC 44121

Qy        350 CGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCT 409

Db      44122 CACGCCTGGCCAG---TTATTTTTTAAAAAGTTTTTGGCCAGGCACGGTGGCTCACGTCT 44178

Qy        410 GTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGA 469

Db      44179 GTAATCCCAGCACTTTGGGAGGACAAGGCAGGTGGATCACTTGAGGTCAGGAGTTTGAGA 44238

Qy        470 CCAGCCAGGCCAACATGGTAAAACCATGTC 499

Db      44239 CCAGTTTGGCCAACATGTTCAAACCCCATC 44268


RESULT 10
AC135837
LOCUS       AC135837              169934 bp    DNA     linear   HTG 18-JAN-2003
DEFINITION  Papio anubis clone RP41-316H21, WORKING DRAFT SEQUENCE, 6 ordered
            pieces.
ACCESSION   AC135837
VERSION     AC135837.3  GI:27802021
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 169934)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169934)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 169934)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jan 18, 2003 this sequence version replaced gi:27261523.
            -------------- Genome Center
                Center: NIH Intramural Sequencing Center
                Center code: NISC
                Web site: http://www.nisc.nih.gov
                Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                Center project name: cqg
                Center clone name: 316H21

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                Sequencing vector: plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.990319
                Consensus quality: 169111 bases at least Q40
                Consensus quality: 169324 bases at least Q30
                Consensus quality: 169383 bases at least Q20
                Insert size: 163000; agarose-fp
                Insert size: 169434; sum-of-contigs
                Quality coverage: 11.81x in Q20 bases; agarose-fp
                Quality coverage: 11.36x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
```

```
         * the accession number will be preserved.
         *        1     24041: contig of 24041 bp in length
         *    24042     24141: gap of unknown length
         *    24142     52259: contig of 28118 bp in length
         *    52260     52359: gap of unknown length
         *    52360     71855: contig of 19496 bp in length
         *    71856     71955: gap of unknown length
         *    71956    134599: contig of 62644 bp in length
         *   134600    134699: gap of unknown length
         *   134700    140490: contig of 5791 bp in length
         *   140491    140590: gap of unknown length
         *   140591    169934: contig of 29344 bp in length.
FEATURES             Location/Qualifiers
     source          1..169934
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-316H21"
                     /clone_lib="RP41"
     misc_feature    1..37142
                     /note="clone overlaps with GenBank Accession Number
                     AC135835 clone RP41-232D1 (center project name cqp)"
     misc_feature    1..24041
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    24142..52259
                     /note="assembly_fragment"
     misc_feature    52360..71855
                     /note="assembly_fragment"
     misc_feature    71956..134599
                     /note="assembly_fragment"
     misc_feature    134700..140490
                     /note="assembly_fragment"
     misc_feature    140066..169934
                     /note="clone overlaps with GenBank Accession Number
                     AC116931 clone RP41-152L1 (center project name cqr)"
     misc_feature    140591..169934
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    45840 a  37767 c  37984 g  47843 t    500 others
ORIGIN

  Query Match             36.3%;  Score 181.4;  DB 2;  Length 169934;
  Best Local Similarity   74.3%;  Pred. No. 2.7e-35;
  Matches  243;  Conservative    0;  Mismatches   81;  Indels    3;  Gaps    1;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232

Db      29933 TTATTTATTTTAATTTTTTTTATTTTATTCTCATTTTTTTGGAGATGGAGTCTCACTCTG 29992

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292

Db      29993 TCACCCAGGCTGCAGTGCAGTGGCATGATCTCGGCTCACTGCAACCTCTGCCTCCTGAGT 30052

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352

Db      30053 TCAAGCGATTCTCCTGCTTCAGGCTCCCAAGTGGCTGGGGCTACAGGCATGCACCACCGC 30112

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412

Db      30113 ACCTGGCTGATTTTTTTATTTTTTA---GTAGAGACAGGTTTCACCATATTTCTCCTGTA 30169

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472

Db      30170 ATCCCAGCACTTTGGGAGGCCAAGGTAGGCAGATCACCTGAGGTCAGGAGTTCCAGACCA 30229

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499

Db      30230 GCCTGACCAACATGGAGAAACCCTGTC 30256
```

```
RESULT 11
AC135835
LOCUS       AC135835              184341 bp    DNA     linear   HTG 25-JAN-2003
DEFINITION  Papio anubis clone RP41-232D1, WORKING DRAFT SEQUENCE, 8 ordered
            pieces.
ACCESSION   AC135835
VERSION     AC135835.3  GI:27901836
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 184341)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 184341)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 184341)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JAN-2003) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Jan 25, 2003 this sequence version replaced gi:27476125.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            -------------- Project Information
                 Center project name: cqp
                 Center clone name: 232D01

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 182494 bases at least Q40
                 Consensus quality: 183101 bases at least Q30
                 Consensus quality: 183419 bases at least Q20
                 Insert size: 154000; agarose-fp
                 Insert size: 183641; sum-of-contigs
                 Quality coverage: 11.91x in Q20 bases; agarose-fp
                 Quality coverage: 9.99x in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
```

```
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1     5018: contig of 5018 bp in length
            *    5019     5118: gap of unknown length
            *    5119    60125: contig of 55007 bp in length
            *   60126    60225: gap of unknown length
            *   60226    67350: contig of 7125 bp in length
            *   67351    67450: gap of unknown length
            *   67451   109121: contig of 41671 bp in length
            *  109122   109221: gap of unknown length
            *  109222   147612: contig of 38391 bp in length
            *  147613   147712: gap of unknown length
            *  147713   150381: contig of 2669 bp in length
            *  150382   150481: gap of unknown length
            *  150482   171241: contig of 20760 bp in length
            *  171242   171341: gap of unknown length
            *  171342   184341: contig of 13000 bp in length.
FEATURES             Location/Qualifiers
     source          1..184341
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-232D1"
                     /clone_lib="RP41"
     misc_feature    1..45592
                     /note="clone overlaps with GenBank Accession Number
                     AC135547 clone RP41-213D9 (center project name cqo)"
     misc_feature    1..5018
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    5119..60125
                     /note="assembly_fragment"
     misc_feature    60226..67350
                     /note="assembly_fragment"
     misc_feature    67451..109121
                     /note="assembly_fragment"
     misc_feature    109222..147612
                     /note="assembly_fragment"
     misc_feature    147169..184341
                     /note="clone overlaps with GenBank Accession Number
                     AC135837 clone RP41-316H21 (center project name cqg)"
     misc_feature    147713..150381
                     /note="assembly_fragment"
     misc_feature    150482..171241
                     /note="assembly_fragment"
     misc_feature    171342..184341
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
BASE COUNT    51268 a  42863 c  43040 g  46468 t    702 others
ORIGIN

  Query Match             36.3%;  Score 191.4;  DB 2;  Length 184341;
  Best Local Similarity   74.3%;  Pred. No. 2.7e-35;
  Matches  243;  Conservative    0;  Mismatches   81;  Indels    3;  Gaps    1;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232
              |  | | | ||  || |||| |||  ||     ||  ||  ||||||| |  ||||
Db     177131 TTATTTATTTAATTTTTTTTATTTTATTCTCATTTTTTTGGAGATGGAGTCTCACTCTG 177190

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
               | |||||| || | ||| |||||||||| ||| ||| ||||     |   |
Db     177191 TCACCCAGGCTGCAGTGCAGTGGCATGATCTCGGCTCACTGCAACCTCTGCCTCCTGAGT 177250

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
              ||||| || |||||| ||  || |||||  |||  | |||| ||  |     ||||
Db     177251 TCAAGCGATTCTCCTGCTTCAGGCTCCCAAGTGGCTGGGGCTACAGGCATGCACCACCGC 177310

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTTAAATTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
              |||||||    |  ||  |||| |    || |   | |            |||
Db     177311 ACCTGGCTGATTTTTTATTTTTTA---GTAGAGACAGGTTTCACCATATTTCGCCTGTA 177367

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
              ||||||| ||| ||| |||| |||  ||| ||| || |||||||||||||  ||||
Db     177368 ATCCCAGCACTTTGGGAGGCCAAGGTAGGCAGATCACCTGAGGTCAGGAGTTCCAGACCA 177427

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
              ||| | |||||||   |||||  ||
Db     177428 GCCTGACCAACATGGAGAAACCCTGTC 177454


RESULT 12
AL929255
LOCUS       AL929255               89319 bp    DNA     linear   HTG 04 OCT-2002
DEFINITION  Homo sapiens chromosome 1 clone RP4-626J17, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   AL929255
VERSION     AL929255.2  GI:23558421
KEYWORDS    HTG; HTGS_PHASE2; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 89319)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-OCT-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 7, 2002 this sequence version replaced gi:23395929.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ626J17
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 88731 bases at least Q40
            Consensus quality: 88854 bases at least Q30
            Consensus quality: 88921 bases at least Q20
            Insert size: 89319; sum-of-contigs
            Insert size: 117285; 11.2% error; agarose-fp
            Quality coverage: 12.74x in Q20 bases; sum-of-contigs Quality
            coverage: 9.71x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..89319
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP4-626J17"
                     /clone_lib="RPCI-4"
     misc_feature    1..89319
                     /note="assembly_fragment:C2502
                     clone_end:SP6
                     vector_side:right
                     clone_end:T7
                     vector_side:left"
BASE COUNT    23401 a  20390 c  21063 g  24465 t
ORIGIN

  Query Match             35.9%;  Score 179.4;  DB 2;  Length 89319;
  Best Local Similarity   75.1%;  Pred. No. 8.5e-35;
```

```
Matches  238;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    1;

Qy        186 TTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGG 245
Db       1129 TTGAAGCAACTTTTTTTTTTTTGTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGCTGG 1188

Qy        246 AATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTC 305
Db       1189 AGTGCAATGGCGTGATCTTGACTCACGGCAACCTCTGCCTCCTGGGTTCGAGCGATTCTC 1248

Qy        306 CTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTT 365
Db       1249 CTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATCCACCACTACGTCCGGCTAATAA 1308

Qy        366 TGTTTGTTTTTAAATT---TTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCAC 422
Db       1309 AGTTGTTTTTAAAAGTTGCTTCAGCCTGGGTGCAGTGGCTCATGCCTGTAACCCCAGCAC 1368

Qy        423 TTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAA 482
Db       1369 TTTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACTAGCCTGGCCAA 1428

Qy        483 CATGGTAAAACCATGTC 499
Db       1429 CATGGCAAAACCCCATC 1445
```

```
RESULT 13
AL607089/c
LOCUS       AL607089                97621 bp    DNA     linear   PRI 14-FEB-2002
DEFINITION  Human DNA sequence from clone RP11-248I9 on chromosome 1, complete
            sequence.
ACCESSION   AL607089
VERSION     AL607089.16  GI:18693046
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97621)
  AUTHORS   Brown,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 15, 2002 this sequence version replaced gi:18476794.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-248I9 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-248I9 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true left end of clone RP11-181E22 is at 95622 in this
            sequence. The true right end of clone RP11-244H3 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..97621
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-248I9"
                     /clone_lib="RPCI-11.1"
     misc_feature    9942..9957
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only"
BASE COUNT    27015 a  20723 c  19956 g  29927 t
ORIGIN

  Query Match             35.9%;  Score 179.4;  DB 9;  Length 97621;
  Best Local Similarity   74.2%;  Pred. No. 8.6e-35;
  Matches  241;  Conservative    0;  Mismatches   81;  Indels    3;  Gaps    1;

Qy        178 TTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCC 237
Db       2829 TTTAACTTTCTTTTTTCTTTTCTTTTTTTTTTGAGGCAGAGGCTCGCTCTGTTGCC 2770

Qy        238 CAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAG 297
Db       2769 CAGGCTGGAGTGCAGTGGCACAATCTTGGCTTACTGCAGCCTCCTCCCGGGTTCAAG 2710

Qy        298 CAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTG 357
Db       2709 TGATTCTCCTGCCTCAGCCTCCTGGGTAGCTGGGACTATAGGTGTGAGCCACCACGCCCA 2650

Qy        358 GCCTGGTTTGTTT---GTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGAT 414
Db       2649 GCCAGATTTAATTTCATTTATTGCTCCTTTTTGCCGGGTGCAGTGGCTCACACCTGTAAT 2590

Qy        415 CCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGC 474
Db       2589 CCCAGCACTTTGGGAGGCCAAAGTGGGCGGATCACCTGAGGTCAGGAGTTCCAGACCAGC 2530

Qy        475 CAGGCCAACATGGTAAAACCATGTC 499
Db       2529 CTGGCTAACATAGCAAAACCCCTTC 2505
```

```
RESULT 14
AL359203/c
LOCUS       AL359203               121362 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Human DNA sequence from clone RP11-547P4 on chromosome 1, complete
            sequence.
ACCESSION   AL359203
VERSION     AL359203.22  GI:21211707
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 121362)
  AUTHORS   Johnson,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:19068224.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
```

```
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-547P4 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..121362
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-547P4"
                     /clone_lib="RPCI-11.2"
BASE COUNT    36463 a  23081 c  24157 g  37661 t
ORIGIN

  Query Match             35.9%;  Score 179.4;  DB 9;  Length 121362;
  Best Local Similarity   76.4%;  Pred. No. 8.6e-35;
  Matches  233;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        195 TTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATG 254
Db      90865 TTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCACTCTGTCGCCCAGGCTGGAGTGCAATG 90806

Qy        255 GCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAA 314
Db      90805 GCATGATCTCGGCTCACTGCAACCTCCGCCTCCCGGGCTCAAGTGATTCTCCTGCCTCAG 90746

Qy        315 TCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTT 374
Db      90745 CCTCCCATGTAGTTGAGATTACAGGCGCCCACCACCACACCTGTGCGTGCCCATATTTTA 90686

Qy        375 TTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCA 434
Db      90685 AGAGATAGT-TGGCCTGGTGTAGTGACTCATGCCTGTAATCCCAGCACTTTGGGAGGCCG 90627

Qy        435 AGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACC 494
Db      90626 AGGCAGGCGGATTACCTGAGGTGAGGAGTTCGAGACCAGCCTGGACAACACGGTGAAACC 90567

Qy        495 ATGTC 499
Db      90566 CAGTC 90562

RESULT 15
AL451139
LOCUS       AL451139               145736 bp    DNA     linear   PRI 22-MAY-2002
DEFINITION  Human DNA sequence from clone RP11-569G9 on chromosome 1, complete
            sequence.
ACCESSION   AL451139
VERSION     AL451139.40  GI:21211745
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 145736)
  AUTHORS   Van Hellmond,Z.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On May 25, 2002 this sequence version replaced gi:20975434.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP11-569G9 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..145736
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-569G9"
                     /clone_lib="RPCI-11.2"
BASE COUNT    36953 a  37564 c  36509 g  34710 t
ORIGIN

  Query Match             35.9%;  Score 179.4;  DB 9;  Length 145736;
  Best Local Similarity   75.1%;  Pred. No. 8.6e-35;
  Matches  238;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    1;

Qy        186 TTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGG 245
Db     122795 TTGAAGCAACTTTTTTTTCTTGTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGCTGG 122854

Qy        246 AATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTC 305
Db     122855 AGTGCAATGGCGTGATCTTGACTCACGGCAACCTCTGCCTCCTGGGTTCGAGCGATTCTC 122914

Qy        306 CTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTT 365
Db     122915 CTGCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATCCACCACTACGTCCGGCTAATAA 122974

Qy        366 TGTTTGTTTTTAAATT---TTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCAC 422
Db     122975 AGTTGTTTTTAAAAGTTGCTTCAGCCTGGGTGCAGTGGCTCATGCCTGTAACCCCAGCAC 123034

Qy        423 TTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAA 482
Db     123035 TTTGGGAGGCCGAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACTAGCCTGGCCAA 123094

Qy        483 CATGGTAAAACCATGTC 499
Db     123095 CATGGCAAAACCCCATC 123111

Search completed: October 24, 2003, 21:05:57
```

Job time : 1386.52 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 125.999 Seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: 1 taactgactctgctgtgtttt.....aacatggtaaaaccatgtcg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1348719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	500	100.0	110000	22	AAF84800	Nucleotide sequenc
2	175.6	35.1	57273	24	ABK22784	Human high bone ma
3	175.6	35.1	66933	22	ABA82625	Human HBM gene reg
4	175.6	35.1	66933	25	ACC45366	Human HBM gene fra
c 5	170.6	34.1	128722	24	ABQ88117	Human osteoblast d
6	169.2	33.8	33147	22	AAK67282	Human immune/haema
c 7	168.6	33.7	314	22	AAK83537	Human immune/haema
c 8	167.8	33.6	51469	22	AAK69322	Human immune/haema
c 9	167.8	33.6	51469	22	AAK70270	Human immune/haema
c 10	167.8	33.6	51469	22	AAK78813	Human immune/haema
c 11	167.2	33.4	2480	22	AAC85484	Human neuropeptide
c 12	166.6	33.3	249999	25	ABZ80229	Human tramdorin ge
c 13	165.8	33.2	32152	22	AAS39621	Genomic sequence #
c 14	165.8	33.2	32152	22	AAK69020	Human digestive sy
15	165.8	33.2	32152	22	AAK91534	Human digestive sy
16	165.8	33.2	32152	22	AAI57791	Human colorectal c
17	165.8	33.2	32152	24	ABS99968	Genomic DNA #172 e
c 18	165.8	33.2	32204	22	AAS39620	Genomic sequence #
c 19	165.8	33.2	32204	22	AAK89018	Human digestive sy
20	165.8	33.2	32204	22	AAK91533	Human digestive sy
21	165.8	33.2	32204	22	AAI57790	Human colorectal c
22	165.8	33.2	32204	24	ABS99967	Genomic DNA #171 e
c 23	165.8	33.2	36933	22	AAK66362	Human immune/haema
24	163.6	32.7	36305	24	ABK22783	Human high bone ma
25	163.6	32.7	72049	22	ABA82623	Human HBM gene reg
26	163.6	32.7	72049	25	ACC45364	Human HBM gene fra
27	163	32.6	9163	24	ABL45808	Human endothelial
28	162.4	32.5	14525	22	AAK70511	Human immune/haema
c 29	162.2	32.4	13024	22	AAL04739	Human reproductive
c 30	162.2	32.4	13024	23	ABL97644	Human testicular a
c 31	162	32.4	4447	22	AAK75249	Human immune/haema
32	160.8	32.2	1331	22	AAK66937	Human immune/haema
c 33	160.4	32.1	3459	22	AAK90901	Human digestive sy
c 34	160.4	32.1	3459	22	AAS31936	Human liver associ
c 35	160.4	32.1	3459	24	ABN90291	Human liver antige
c 36	160	32.0	84607	20	AAX90947	Human PACAP genomi
37	157.8	31.6	8418	22	AAS41906	Genomic sequence #
38	157.6	31.5	1328	22	AAK66936	Human immune/haema
c 39	157	31.4	318	22	AAK62834	Human immune/haema
40	156.8	31.4	24606	22	AAK78367	Human immune/haema
c 41	156.8	31.4	24606	22	AAK84723	Human immune/haema
c 42	156.8	31.4	31051	22	AAK73223	Human immune/haema
43	156.2	31.2	1043	20	AAZ24882	Human secreted pro
c 44	155	31.0	10012	24	ABL55889	Human small induci
c 45	155	31.0	11096	22	ABA18278	Human nervous syst

ALIGNMENTS

RESULT 1
AAF84800
ID AAF84800 standard; DNA; 110000 BP.
XX
AC AAF84800;
XX
DT 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 gene.
XX
KW Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW autosomal dominant familial spastic paraplegia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 9932..102009
FT /*tag= a
FT /note= "contains introns"
FT exon 9932..10471
FT /*tag= b
FT /number= "1"
FT intron 10472..33718
FT /*tag= c
FT /number= "1"
FT exon 33719..33805
FT /*tag= d
FT /number= "2"
FT intron 33806..35748
FT /*tag= e
FT /number= "2"

```
FT   exon            35749..35832
FT                   /*tag= f
FT                   /number= "3"
FT   intron          35833..45022
FT                   /*tag= g
FT                   /number= "3"
FT   exon            45023..45118
FT                   /*tag= h
FT                   /number= "4"
FT   intron          45119..60863
FT                   /*tag= i
FT                   /number= "4"
FT   exon            60864..61051
FT                   /*tag= j
FT                   /number= "5"
FT   intron          61052..61927
FT                   /*tag= k
FT                   /number= "5"
FT   exon            61928..62061
FT                   /*tag= l
FT                   /number= "6"
FT   intron          62062..62344
FT                   /*tag= m
FT                   /number= "6"
FT   exon            62345..62438
FT                   /*tag= n
FT                   /number= "7"
FT   intron          62439..73173
FT                   /*tag= o
FT                   /number= "7"
FT   exon            73174..73248
FT                   /*tag= p
FT                   /number= "8"
FT   intron          73249..74633
FT                   /*tag= q
FT                   /number= "8"
FT   exon            74634..74705
FT                   /*tag= r
FT                   /number= "9"
FT   intron          74706..82788
FT                   /*tag= s
FT                   /number= "9"
FT   exon            82789..82964
FT                   /*tag= t
FT                   /number= "10"
FT   intron          82965..83102
FT                   /*tag= u
FT                   /number= "10"
FT   exon            83103..83194
FT                   /*tag= v
FT                   /number= "11"
FT   intron          83195..83334
FT                   /*tag= w
FT                   /number= "11"
FT   exon            83335..83414
FT                   /*tag= x
FT                   /number= "12"
FT   intron          83415..88129
FT                   /*tag= y
FT                   /number= "12"
FT   exon            88130..88172
FT                   /*tag= z
FT                   /number= "13"
FT   intron          88173..89561
FT                   /*tag= aa
FT                   /number= "13"
FT   exon            89562..89641
FT                   /*tag= ab
FT                   /number= "14"
FT   intron          89642..91162
FT                   /*tag= ac
FT                   /number= "14"
FT   exon            91163..91233
FT                   /*tag= ad
FT                   /number= "15"
FT   intron          91234..93443
FT                   /*tag= ae
FT                   /number= "15"
FT   exon            93444..93484
FT                   /*tag= af
FT                   /number= "16"
FT   intron          93485..100599
FT                   /*tag= ag
FT                   /number= "16"
FT   exon            100600..102009
FT                   /*tag= ah
FT                   /number= "17"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening  -
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 other;

  Query Match             100.0%;  Score 500;  DB 22;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 1.6e-94;
  Matches  500;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TAACTGACTCTGCTGTGTTTCCTTGGCATTATAGCTAATCAAATTGAGCAGGTCAGGTAA 60

Qy         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAGTTTATACTTACACCTACTATTTCAAAACCATGAGCTCATTCACATTTTCACTGAAGT 120

Qy        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTT 180

Qy        181 TTTTGTTGTTTTATTTATTTATTTATTTATTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTGTTGTTTTATTTATTTATTTATTTATTTCTAGATGGAGTCTTGCTCTGTTGCCCAG 240

Qy        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAA 300

Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420

Db        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480

Db        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480

Qy        481 AACATGGTAAAACCATGTCG 500

Db        481 AACATGGTAAAACCATGTCG 500


RESULT 2
ABK22784
ID   ABK22784 standard; cDNA; 57273 BP.
XX
AC   ABK22784;
XX
DT   09-APR-2002  (first entry)
XX
DE   Human high bone mass (HBM) polynucleotide clone #7.
XX
KW   Human; mouse; Zmax1; HBM; high bone mass gene; lipid regulation; stroke;
KW   lipid-associated condition; arteriosclerosis; cardiovascular disease; ss;
KW   osteoporosis; atherosclerosis; diabetic atherosclerosis; plaque build-up;
KW   neurovascular condition; wound healing; gene therapy; PCR primer; probe;
KW   bone development disorder; antiarteriosclerotic; cardiovascular;
KW   osteopathic; cerebroprotective.
XX
OS   Homo sapiens.
XX
PN   WO200192891-A2.
XX
PD   06-DEC-2001.
XX
PF   25-MAY-2001; 2001WO-US16946.
XX
PR   26-MAY-2000; 2000US-0578900.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
PA   (UYCR-) UNIV CREIGHTON SCHOOL MEDICINE.
XX
PI   Carulli JP,  Little RD,  Recker RR,  Johnson ML;
XX
DR   WPI; 2002-097784/13.
XX
PT   Identifying molecules involved in lipid regulation, useful for
PT   diagnosing, treating or preventing e.g., arteriosclerosis, comprises
PT   identifying a molecule that binds to high bone mass gene or its
PT   corresponding wild type gene -
XX
PS   Example 2; Page 350-392; 409pp; English.
XX
CC   The invention relates to a method for identifying a molecule involved in
CC   lipid regulation comprising identifying a molecule that binds to or
CC   inhibits binding of a molecule to high bone mass (HBM) or its wild type
CC   gene, Zmax1. Compounds identified by the method are useful for treating,
CC   diagnosing, preventing or screening for normal and abnormal
CC   lipid-associated conditions, including arteriosclerosis, cardiovascular
CC   disease, stroke, and osteoporosis. The compounds may also be used in the
CC   treatment or prevention of diabetic atherosclerosis, neurovascular
CC   conditions caused by plaque build-up, poor circulation due to plaque
CC   build-up and associated poor wound healing. The methods may be used in
CC   gene therapy, pharmaceutical development, and diagnostic assays for bone
CC   development disorders. Molecules identified by comparison of Zmax1 and
CC   HBM systems can be used as surrogate markers in pharmaceutical
CC   development, in diagnosis of human or animal bone disease, and in the
CC   treatment of bone diseases. Sequences ABK22776-ABK23411 represent cDNA
CC   molecules encoding human Zmax1 and HBM, and PCR primers, probes, linkers
CC   and adapters of the invention.
XX
SQ   Sequence 57273 BP; 12179 A; 15326 C; 15607 G; 14161 T; 0 other;

  Query Match             35.1%;  Score 175.6;  DB 24;  Length 57273;
  Best Local Similarity   75.9%;  Pred. No. 2.4e-27;
  Matches  243;  Conservative    0;  Mismatches   74;  Indels    3;  Gaps    2;

Qy        182 TTTGTTGTTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGG 241

Db      25334 CTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCA-CCAGC 25392

Qy        242 CTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAAT 301

Db      25393 CTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTTTGCCTCCTATGTTCAAGTGAT 25452

Qy        302 CCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCT 361

Db      25453 TTTCCTGCCTCAGCCTCCTGAGTAACTGGGATTACAGGCATGAGCCACCGCGCGGCCGGCCA 25512

Qy        362 GGTTTGTTTGTTTTTA--AATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAG 419

Db      25513 GATTTGCATTTTTGAAACAACTGCTAGGCTGGGCGCGGTGGCTCACACCTGTAATCCCAG 25572

Qy        420 CACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGC 479

Db      25573 CACTGTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGGGTTCGAGACCAGCCTGGC 25632

Qy        480 CAACATGGTAAAACCATGTC 499

Db      25633 CAACATGGTGAAACCCCGTC 25652


RESULT 3
ABA82625
ID   ABA82625 standard; DNA; 66933 BP.
XX
AC   ABA82625;
XX
DT   25-JAN-2002  (first entry)
XX
DE   Human HBM gene region b200e21-h_contig4.
XX
KW   Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
KW   sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
KW   antisense therapy; vaccine; bone disorder; Paget's disease;
KW   sclerostosis; osteomalacia; fibrous dysplasia; ds.
XX
OS   Homo sapiens.
XX
PN   WO200177327-A1.
XX
PD   18-OCT-2001.
XX
PF   21-JUN-2000; 2000WO-US16951.
XX
PR   05-APR-2000; 2000US-0543771.
PR   05-APR-2000; 2000US-0544398.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
XX
PI   Carulli JP,  Little RD,  Recker RR,  Johnson ML;
XX
DR   WPI; 2001-657171/75.
XX
PT   New high bone mass (HBM) and Zmax1 genes and proteins useful for
PT   modulating bone mass for the treatment of e.g. osteoporosis -
XX
PS   Claim 51; Page 308-350; 443pp; English.
XX
CC   The present invention describes the human Zmax1 gene and the high bone
CC   mass (HBM) gene,which are found on chromosome 11q13.3. The Zmax1 and
CC   HBM genes have osteopathic activities. The genes can be used in gene
CC   therapy, antisense therapy and in the production of vaccines. They
```

```
CC   can be used in the diagnosis and treatment of bone disorders including
CC   osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
CC   dysplasia. ABA82036 to ABA82700 and AAG68168 to AAG68193 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;

  Query Match           35.1%;  Score 175.6;  DB 22;  Length 66933;
  Best Local Similarity   75.9%;  Pred. No. 2.5e-27;
  Matches  243;  Conservative    0;  Mismatches   74;  Indels    3;  Gaps    2;

Qy        182 TTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGG 241
Db      34994 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCA-CCAGC 35052

Qy        242 CTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAAT 301
Db      35053 CTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTTTGCCTCCTATGTTCAAGTGAT 35112

Qy        302 CCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCT 361
Db      35113 TTTCCTGCCTCAGCCTCCTGAGTAACTGGGATTACAGGCATGAGCCACCGCGGCCGGCCA 35172

Qy        362 GGTTTGTTTGTTTTTA--AATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAG 419
Db      35173 GATTTGCATTTTTGAAACAACTGCTAGGCTGGGCGCGGTGGCTCACACCTGTAATCCCAG 35232

Qy        420 CACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGC 479
Db      35233 CACTGTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGGGTTCGAGACCAGCCTGGC 35292

Qy        480 CAACATGGTAAAACCATGTC 499
Db      35293 CAACATGGTGAAACCCCGTC 35312


RESULT 4
ACC45366
ID   ACC45366 standard; DNA; 66933 BP.
XX
AC   ACC45366;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human HBM gene fragment #7.
XX
KW   Human; high bone mass; HBM; LRP5; LRP6; transgenic; bone mass modulation;
KW   gene therapy; bone density modulation; bone strength; trabecular number;
KW   bone size; bone tissue connectivity; bone disease; osteoporosis;
KW   osteomalacia; rickets; Paget's disease; neoplasm of the bone; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200292764-A2.
XX
PD   21-NOV-2002.
XX
PF   13-MAY-2002; 2002WO-US14876.
XX
PR   11-MAY-2001; 2001US-290071P.
PR   17-MAY-2001; 2001US-291311P.
PR   01-FEB-2002; 2002US-353056P.
PR   04-MAR-2002; 2002US-361293P.
XX
PA   (GENO-) GENOME THERAPEUTICS CORP.
PA   (AMHP ) WYETH.
XX
PI   Babij P, Bex FJ, Yaworsky PJ, Bodine PV;
XX
DR   WPI; 2003-129278/12.
XX
PT   New transgenic animals (e.g. mice), useful as models for studying bone
PT   density modulation, developing drugs for treating or preventing bone
PT   diseases (e.g. osteoporosis), or diagnosing diseases characterized by
PT   reduced bone density -
XX
PS   Example 2; Page 362-384; 603pp; English.
XX
CC   The invention relates to novel transgenic animals expressing the high
CC   bone mass (HBM) gene, expressing the corresponding wild type HBM gene,
CC   comprising an alteration of the gene encoding LRP5 or LRP6, or
CC   expressing an LRP5 that is modulated by an altered gene control
CC   sequence introduced by homologous or non-homologous recombination. The
CC   transgenic animals are for the study of bone density modulation or bone
CC   mass modulation. The invention has osteopathic and cytostatic activity.
CC   The polynucleotides of the invention may have a use in gene therapy.
CC   The transgenic animals and nucleic acids are for the study of
CC   bone density modulation, where the bone mass is modulated relative to
CC   non-transgenic animals of the same species in more than one parameter
CC   selected from bone density, bone strength, trabecular number, bone
CC   size, or bone tissue connectivity. The transgenic animals, nucleic
CC   acids and methods are useful for identifying molecules involved in bone
CC   development, and for developing pharmaceutical compositions, which may
CC   be employed for treating or preventing bone diseases, e.g.
CC   osteoporosis, osteomalacia, rickets, Paget's disease, or neoplasms of
CC   the bone. The transgenic animals and nucleic acids are also useful in
CC   methods for diagnosing diseases involved in bone development, or
CC   characterised by reduced bone density or mass. The present sequence is
CC   used in the exemplification of the invention.
XX
SQ   Sequence 66933 BP; 14237 A; 17817 C; 18323 G; 16556 T; 0 other;

  Query Match           35.1%;  Score 175.6;  DB 25;  Length 66933;
  Best Local Similarity   75.9%;  Pred. No. 2.5e-27;
  Matches  243;  Conservative    0;  Mismatches   74;  Indels    3;  Gaps    2;

Qy        182 TTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGG 241
Db      34994 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCA-CCAGC 35052

Qy        242 CTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAAT 301
Db      35053 CTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTTTGCCTCCTATGTTCAAGTGAT 35112

Qy        302 CCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCT 361
Db      35113 TTTCCTGCCTCAGCCTCCTGAGTAACTGGGATTACAGGCATGAGCCACCGCGGCCGGCCA 35172

Qy        362 GGTTTGTTTGTTTTTA--AATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAG 419
Db      35173 GATTTGCATTTTTGAAACAACTGCTAGGCTGGGCGCGGTGGCTCACACCTGTAATCCCAG 35232

Qy        420 CACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGC 479
Db      35233 CACTGTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGGGTTCGAGACCAGCCTGGC 35292

Qy        480 CAACATGGTAAAACCATGTC 499
Db      35293 CAACATGGTGAAACCCCGTC 35312


RESULT 5
ABQ88117/c
ID   ABQ88117 standard; cDNA; 129722 BP.
XX
AC   ABQ88117;
XX
DT   18-SEP-2002  (first entry)
XX
DE   Human osteoblast differentiation related cDNA SEQ ID NO 24.
XX
KW   Human; osteoblast; stem cell differentiation; bone tissue deposition;
KW   osteoporosis; osteopathic; ss.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200250301-A2.
XX
PD   27-JUN-2002.
XX
PF   18-DEC-2001; 2001WO-US48276.
XX
PR   18-DEC-2000; 2000US-255882P.
PR   24-APR-2001; 2001US-285691P.
XX
PA   (GENE-) GENE LOGIC INC.
PA   (PROC ) PROCTER & GAMBLE CO.
XX
PI   Ji D,  Axelrod DW,  Cook JS,  Jaiswal N,  Einstein R,  Houghton A;
PI   Mertz L;
XX
DR   WPI; 2002-557663/59.
XX
PT   Use of genes and their expression profiles associated with osteoblast
PT   differentiation for screening modulators bone formation, for diagnosing
PT   or treating e.g. osteoporosis, or as markers for the differentiation
PT   process  -
XX
PS   Claim 1; SEQ ID NO 24; 78pp + Sequence Listing; English.
XX
CC   The invention relates to genes and their expression profiles are used
CC   for:
CC   (a) screening modulators of precursor stem cell differentiation into
CC   osteoblasts, or bone tissue deposition;
CC   (b) diagnosing abnormal deposition of bone tissue, abnormal rate of
CC   osteoblast formation or osteoporosis; or
CC   (c) treating or monitoring treatment of the conditions cited in (b), or
CC   monitoring the progression of bone tissue deposition.
CC   Specific conditions include postmenopausal osteoporosis, glucocorticoid
CC   osteoporosis or male osteoporosis, osteopenia, osteodystrophy,
CC   drug-induced abnormalities in bone formation or bone loss, conditions
CC   that involve altered bone metabolism (e.g. idiopathic juvenile
CC   osteoporosis), skeletal disease linked to breast cancer, mastocytosis,
CC   Fanconi syndrome or fibrous dysplasia. The present sequence is that of an
CC   osteoblast differentiation associated cDNA marker of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 129722 BP; 33298 A; 32697 C; 31415 G; 32312 T; 0 other;

  Query Match             34.1%;  Score 170.6;  DB 24;  Length 129722;
  Best Local Similarity   72.3%;  Pred. No. 2.8e-26;
  Matches  235;  Conservative    0;  Mismatches   89;  Indels    1;  Gaps    1;

Qy        176 TGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTG 235

Db      14463 TGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAAGACAGTGTCTCATTCTGTCA 14404

Qy        236 CCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCA 295

Db      14403 CTCAGGCTGGAGTGCAGTGGCATGATTATAGCTTACTGCAACCTCGAACTCCTGGGCTCA 14344

Qy        296 AGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACC 355

Db      14343 AGGGCTCCTCTCATCCCAGCCTCCCAAATAGCTGGGACTACAGGTGTGTGTCACAGCACT 14284

Qy        356 TGGCCTGGTTTGTTTGTTTTTAAATTTTGA-GGCCAGGTGCAGTGGCCCATATCTGTGAT 414

Db      14283 TGGCTAACTTTTTTAAAAAAATTTTTTGTAATGGTCGGGTGCAGTGCCTCATGCCTGTAAT 14224

Qy        415 CCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGC 474

Db      14223 CCCAGCACTTTGGGAGGCCAAGACAGGAGGATCAGTTGAGGTCAGGAGTTTGAAACCAGC 14164

Qy        475 CAGGCCAACATGGTAAAACCATGTC 499

Db      14163 TTGGCCAACATGGCGAAACCCCATC 14139
```

```
RESULT 6
AAK67282
ID   AAK67282 standard; DNA; 33147 BP.
XX
AC   AAK67282;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:22094.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   21-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
```

PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
PS Disclosure; SEQ ID NO 22094; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 33147 BP; 8370 A; 7682 C; 8382 G; 8713 T; 0 other;

Query Match 33.8%; Score 169.2; DB 22; Length 33147;
Best Local Similarity 70.3%; Pred. No. 5e-26;
Matches 242; Conservative 0; Mismatches 98; Indels 4; Gaps 1;

Qy 156 GTGACTGGGAGTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTA 215
Db 17035 GTGACATGATATCTACAGCTAACTTTTATTTCACTTATTTATTATTATTATTATTTTAGA 17094

Qy 216 GATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCA 275
Db 17095 GACAGGGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGTGGCATGATCGTAGCTCACTGCA 17154

Qy 276 ACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTA 335
Db 17155 GCCTCAAACTCCTGAGATCAAGCGATCCTCCTGCCTCAGCCTCTCCAATAGCTGGGACTA 17214

Qy 336 CAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGC 395
Db 17215 CAGGCCCACACCACCATGCCTGGCTTACAACTAACTTTTAAATGTT----GGCCAGGCGC 17270

Qy 396 AGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGG 455
Db 17271 AGTGGCTCACGCCTGAAATCCAAGCACTTTGGGAGGCCGAGGCAGGCGGATCACTTGAGG 17330

Qy 456 TCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
Db 17331 TCAGGAGTTCAAGACCAGTCTCGCCAACATGATGAAACCTCGTC 17374

```
RESULT 7
AAK83537/c
ID   AAK83537 standard; DNA; 314 BP.
XX
AC   AAK83537;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38349.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234225.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234937.
PR   25-SEP-2000; 2000US-0234938.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
```

```
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 38349; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 314 BP; 86 A; 74 C; 87 G; 67 T; 0 other;

  Query Match             35.7%;  Score 169.6;  DB 22;  Length 314;
  Best Local Similarity   74.8%;  Pred. No. 4.6e-26;
  Matches  225;  Conservative    0;  Mismatches   74;  Indels    2;  Gaps    1;

Qy        199 TTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCAT 258
              ||||  :  ||     | :  |||     | : |     ||    ||  :   |   | :|
Db        301 TTTTTTTTCTTTTTTGAAACAAAGTGTCACTCTGCCACCCAGGCTGGGGTGCAGTGGCAC 242

Qy        259 GATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTT 318
               ||||| : ||||  ||      | | :      |||  :      :|   ||| :||    |
Db        241 CATCTCAGCTCACTGCAA--TCTGCCTTCTAGGTTCAAGTGATTCTCCTGCCTCAGCTTC 184

Qy        319 CCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAA 378
              |||| ::||||| :| | |:| |  |     ||   |    :  ||      :|    |: : |
Db        183 CCAAGTAGCTGGGACCACAGGCACATGCCACCATGCCCGGCTAATTTTTGTATTTTTAAT 124

Qy        379 ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGC 438
               :   |   ||::  ||     ||  :  || | |||: |||:|||| ||  |||  || ||:
Db        123 AATTGTAGGCCTGGCACGGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGC 64

Qy        439 AGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGT 498
               :|  ||     ||||:||||| |||||. |||:|||. |||  ||    ||||||!  |
Db         63 GGGTGGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACGTGGTAAAACCATCT 4

Qy        499 C 499
              :
Db          3 C 3
```

```
RESULT 8
AAK69322/c
ID   AAK69322 standard; DNA; 51469 BP.
XX
AC   AAK69322;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24134.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216680.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
```

```
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241797.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 24134; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91321. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 51469 BP; 13004 A; 12589 C; 13305 G; 12571 T; 0 other;

  Query Match             33.6%;  Score 167.8;  DB 22;  Length 51469;
  Best Local Similarity   73.6%;  Pred. No. 1e-25;
  Matches  229;  Conservative    0;  Mismatches   77;  Indels    5;  Gaps    1;

Qy        194 TTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253
              |   |  ||| ||  | |||    ||| | | ||||||  |||||||  ||||||  |||  |
Db      16467 TTTTTCTTTTTTTTTTTTTGAGACAGGGTCTTGCTGTGTTGCCCAGGCTGGAGTGCAGT 16408

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313
              ||  | |||  | |||  | || ||| | ||      | |     ||||| ||||||
Db      16407 GGTGTGATCATAGCTCACTACAGTCTCAAACTCCTAGGCTGAAGAGATCCTCCTGCCTCA 16348

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTT 373
                 | |||||||||||||||||   || |||  |    |   |||| | |    | | ||
Db      16347 GCATCCCAAGTAGCTGGGACTACTGGTATGCACCACCGTGCCTGGCTTATATAGCTTCTC 16288

Qy        374 TTTAA-----ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGG 428
               ||||      ||     |||   | |  ||||| |||  ||||   ||||||| ||||||
Db      16287 CTTAAGAACTCTTCACTGGCTGTGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 16228

Qy        429 AGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGT 488
              ||  |||||||| || ||||||  |||||||||||||   ||||||| | | |    |||||
Db      16227 AGGCCAAGGCAAGCAGATTACCTGAGGTCAGGAGTTTGAGACCAGCCTGACCAGCATGGT 16168

Qy        489 AAAACCATGTC 499
               ||||   |
Db      16167 GAAACCCTATC 16157
```

```
RESULT 9
AAK70270/c
ID   AAK70270 standard; DNA; 51469 BP.
XX
AC   AAK70270;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25082.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227309.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
```

```
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 25082; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 51469 BP; 13004 A; 12589 C; 13305 G; 12571 T; 0 other;

  Query Match             33.6%;  Score 167.8;  DB 22;  Length 51469;
  Best Local Similarity   73.6%;  Pred. No. 1e-25;
  Matches  229;  Conservative    0;  Mismatches   77;  Indels    5;  Gaps    1;

Qy        194 TTTATTTTATTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253
              |||  ||| |||  |||   |   |  |||||  ||| ||||||| ||||||||  ||||
Db      16467 TTTTCTTTTTTTTTTTTTTGAGACAGGGTCTTGCTGTGTTGCCCAGGCTGGAGTGCAGT 16408

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313
              ||  ||||| ||||||||  ||  ||  |     || ||  ||  ||| ||||||||
Db      16407 GGTGTGATCATAGCTCACTACAGTCTCAAACTCCTAGGCTGAAGAGATCCTCCTGCCTCA 16348

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTT 373
                | ||||||||||||||  || ||   ||  ||  ||  |||| |  |  ||
Db      16347 GCATCCCAAGTAGCTGGGACTACTGGTATGCACCACCGTGCCTGGCTTATATAGCTTCTC 16288

Qy        374 TTTAA-----ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGG 428
              ||||      ||   |||   |  |||| |  |||  |||||||||||||
Db      16287 CTTAAGAACTCTTCACTGGCTGTGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 16228

Qy        429 AGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGT 488
              || |||||||  | |||  || || ||||||||  |||    | |||  | |  ||  |||
Db      16227 AGGCCAAGGCAAGCAGATTACCTGAGGTCAGGAGTTTGAGACCAGCCTGACCAGCATGGT 16168

Qy        489 AAAACCATGTC 499
              ||||| | ||
Db      16167 GAAACCCTATC 16157
```

```
RESULT 10
AAK78813/c
ID   AAK78813 standard; DNA; 51469 BP.
XX
AC   AAK78813;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:33625.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
```

```
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 33625; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 51469 BP; 13004 A; 12589 C; 13305 G; 12571 T; 0 other;

  Query Match             33.6%;  Score 167.8;  DB 22;  Length 51469;
  Best Local Similarity   73.6%;  Pred. No. 1e-25;
  Matches  229;  Conservative    0;  Mismatches   77;  Indels    5;  Gaps    1;

Qy        194 TTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253

Db      16467 TTTTTCTTTTTTTTTTTTTTGAGACAGGGTCTTGCTGTGTTGCCCAGGCTGGAGTGCAGT 16408

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313

Db      16407 GGTGTGATCATAGCTCACTACAGTCTCAAACTCCTAGGCTGAAGAGATCCTCCTGCCTCA 16348

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTT 373

Db      16347 GCATCCCAAGTAGCTGGGACTACTGGTATGCACCACCGTGCCTGGCTTATATAGCTTCTC 16288

Qy        374 TTTAA-----ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGG 428

Db      16287 CTTAAGAACTCTTCACTGGCTGTGCGTGGTGGCTCATGCCTGTAATCCCAGCACTTTGGG 16228

Qy        429 AGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGT 488

Db      16227 AGGCCAAGGCAAGCAGATTACCTGAGGTCAGGAGTTTGAGACCAGCCTGACCAGCATGGT 16168

Qy        489 AAAACCATGTC 499

Db      16167 GAAACCCTATC 16157
```

```
RESULT 11
AAC85484/c
ID   AAC85484 standard; cDNA; 2480 BP.
XX
AC   AAC85484;
XX
DT   16-MAY-2001  (first entry)
XX
DE   Human neuropeptide FF promoter.
XX
KW   Promoter; murine; neuropeptide FF; NPFF; brain; spinal cord; alacrima;
KW   inflammation; transcription factor; NFkappaB; hormonal dysfunction;
KW   nuclear factor of activated T-cells; NFAT; heat shock factor 1; ACTH;
KW   HSF1; Allgrove syndrome; triple-A syndrome; sensory impairment;
KW   adrenocorticotropic hormone; resistant adrenal insufficiency;
KW   achalasia; hypoglycaemia; autonomic neuropathy; autonomic function;
KW   gene therapy; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   protein_bind    2102..2113
FT                   /*tag= a
FT                   /bound_moiety= "AP1"
FT   protein_bind    complement (2122..2128)
FT                   /*tag= b
FT                   /bound_moiety= "AP1"
FT   protein_bind    2155..2171
FT                   /*tag= c
FT                   /bound_moiety= "CMYB"
FT   protein_bind    2206..2216
FT                   /*tag= d
FT                   /bound_moiety= "AP1"
FT   protein_bind    2343..2352
FT                   /*tag= e
FT                   /bound_moiety= "NFkappaB"
FT   protein_bind    complement (2373..2385)
FT                   /*tag= f
FT                   /bound_moiety= "NFAT"
FT   protein_bind    complement (2393..2402)
FT                   /*tag= g
FT                   /bound_moiety= "NFAT"
FT   TATA_signal     2441..2445
FT                   /*tag= h
XX
PN   EP1074619-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-0660130.
XX
PR   03-AUG-1999;   99US-0365755.
PR   27-MAR-2000; 2000US-0534638.
XX
PA   (PANU/) PANULA P A J.
PA   (BRAN/) BRANDT A.
PA   (WEST/) WESTERLUND J.
XX
PI   Panula PAJ,  Brandt A,  Westerlund J;
XX
DR   WPI; 2001-171047/18.
XX
PT   New promoter for neuropeptide FF (NPFF), useful for treating and
PT   screening genetic diseases associated with NPFF gene promoter such as
PT   deficient regulation of autonomic function and pain conditions -
XX
PS   Claim 2; Fig 4; 55pp; English.
XX
CC   This sequence represents the promoter region for the human neuro-
CC   peptide FF (NPFF) gene.  The NPFF gene is expressed in specific regions
CC   of the brain and in the spinal cord and is induced upon inflammatory
CC   stimulus.  Binding sites were found for the inflammation related
CC   transcription factor, NFkappaB, and for the nuclear factor of
CC   activated T-cells (NFAT).  A binding site was also found for heat
CC   shock factor 1 (HSF1) which is activated in cells at elevated
CC   temperatures and other environmental stress conditions.  The
CC   AC-dinucleotide repeat is thought to add additional regulatory
CC   effects.  The human NPFF gene is located in the human chromosome
CC   locus 12q13 which is known to be involved in Allgrove syndrome
CC   (triple-A syndrome) which is characterised by a triad of adreno
CC   corticotropic hormone (ACTH), resistant adrenal insufficiency,
CC   achalasia and alacrima, hypoglycaemia and sensory impairment and
CC   autonomic neuropathy.  The NPFF promoter region may be useful for
CC   treating genetic diseases such as those associated with deficient
CC   regulation of autonomic function, pain conditions or hormonal
CC   dysfunction which are associated with the promoter area of the
CC   NPFF gene whose expression is modulated through the regulatory sites.
CC   It is also useful in the screening of the genetic diseases associated
CC   with the promoter area of the NPFF gene by modulation of activation or
CC   inhibition of NPFF gene expression through the regulation sites in the
CC   promoter area, and is used in gene therapy and DNA analysis.
XX
SQ   Sequence 2480 BP; 638 A; 596 C; 673 G; 573 T; 0 other;

  Query Match             33.4%;  Score 167.2;  DB 22;  Length 2480;
  Best Local Similarity   75.6%;  Pred. No. 1.1e-25;
  Matches  239;  Conservative    0;  Mismatches   63;  Indels   14;  Gaps    2;

Qy        193 ATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTC-TGTTGCCCAGGCTGGAATGCA 251
Db       1885 AATTTAACTTTTTTTCTTTTTGAGATGGAGTTTCGCTCTTTTTGCCCAGGCTGGAGTGCA 1826

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
Db       1825 GTGGCGCGATCTCGGCTCACTGCAACCTCTGGCTCCTGGGTTCAAGTGATTCTCCTGCCT 1766

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
Db       1765 CAGCCTCCCAAGTAGCTGGGACTACAGGCGTCAACCACCATGCCCAGCCTGTATCTTTTT 1706

Qy        372 TTTTTAAA-------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCA 418
Db       1705 TTTTTTAATTAAAAACCATTTTTTTTTGCCGGGTGCAGTGCCTCACACCTGTAACCCCA 1646

Qy        419 GCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGG 478
Db       1645 GCACTTTGGGAGACCAAGGCGGGTGGATCACCTGAGGCAGGAGTTCGAGACCAGCCTGG 1586

Qy        479 CCAACATGGTAAAACC 494
Db       1585 CCAACATGGTGAAACC 1570


RESULT 12
ABZ80229/c
ID   ABZ80229 standard; DNA; 249999 BP.
XX
AC   ABZ80229;
XX
DT   02-JUN-2003  (first entry)
XX
DE   Human tramdorin gene region genomic DNA SEQ ID NO:26.
XX
KW   Neuroprotective; nootropic; cerebroprotective; analgesic; gene therapy;
KW   central nervous system disorder; CNS disorder; multiple sclerosis;
KW   nerve injury; neuropathic pain; stroke; trauma; non-CNS disorder; tramd;
KW   tramdorin; human; chromosome 5; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003016502-A2.
XX
PD   27-FEB-2003.
XX
PF   21-AUG-2002; 2002WO-US26637.
```

```
XX
PR   21-AUG-2001; 2001US-313907P.
PR   21-AUG-2002; 2002US-225810.
XX
PA   (MCLA-) MCLAUGHLIN RES INST.
XX
PI   Bermingham JR;
XX
DR   WPI; 2003-278567/27.
XX
PT   New nucleic acid sequence encoding tramdorins, e.g. mouse tramd 1,
PT   mouse tramd 2, mouse tramd 3, human tramd 1, human tramd 2, human tramd
PT   3 or rat tramd 1, useful for treating CNS, e.g. stroke, multiple
PT   sclerosis, trauma, neuropathic pain  -
XX
PS   Example 6; Fig 9; 177pp; English.
XX
CC   The present invention describes an isolated nucleic acid sequence
CC   comprising a cDNA sequence encoding mouse tramdorin (tramd) 2, mouse
CC   tramd 3, human tramd 1, human tramd 2, human tramd 3 or rat tramd 1, or
CC   the genomic sequence of mouse tramd 1 or mouse tramd 3. Mouse tramd 1 is
CC   located to chromosome 11, whereas human tramd 1 is located to chromosome
CC   5q31-33. The tramd sequences have neuroprotective, nootropic, analgesic
CC   and cerebroprotective activities, and can be used in gene therapy. The
CC   nucleic acid sequences are useful for diagnosing and treating central
CC   nervous system (CNS) disorders such as multiple sclerosis, nerve injury,
CC   neuropathic pain, stroke or trauma, and non-CNS disorders. The present
CC   sequence represents the genomic sequence of the human tramdorin gene
CC   region, which is given in the exemplification of the present invention.
XX
SQ   Sequence 249999 BP; 75050 A; 54012 C; 51931 G; 67805 T; 1201 other;

  Query Match             33.3%;  Score 166.6;  DB 25;  Length 249999;
  Best Local Similarity   71.7%;  Pred. No. 2e-25;
  Matches  236;  Conservative    0;  Mismatches   84;  Indels    9;  Gaps    1;

Qy        180 TTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCA 239

Db     211199 TTATGCATTTATTATTTTATTTTATTTTTATTTTTTGAGACAGGGTCTCACTGTCACCCA 211140

Qy        240 GGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCA 299

Db     211139 GGCTGAAGTGCAGTGGCATAATCACAGTTCACTGTGACCTCCTCCTCCCAGGCTCAAGCA 211080

Qy        300 ATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGC 359

Db     211079 ATCCTCCAACCTCAGCCTCCTGAGTAGTTGGGACTACAGGCATTCACCATCATGCCCAGC 211020

Qy        360 CTGGTTTGTTTGTTTTTAAA---------TTTTGAGGCCAGGTGCAGTGGCCCATATCTG 410

Db     211019 TAATTTTTGTATTTCTTGTAGATATGGGGTTTTGTGGCCGGGTGCAGTGGCTCATGCCTG 210960

Qy        411 TGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGAC 470

Db     210959 TAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTAGAGAC 210900

Qy        471 CAGCCAGGCCAACATGGTAAAACCATGTC 499

Db     210899 CAGCCTGGCCAACATGGTGAAACCCCATC 210871


RESULT 13
AAS39621/c
ID   AAS39621 standard; DNA; 32152 BP.
XX
AC   AAS39621;
XX
DT   17-DEC-2001  (first entry)
XX
DE   Genomic sequence #40 encoding human colon associated polypeptide.
XX
KW   Human; colon cancer; congenital abnormality; infection; colitis;
KW   inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW   intestinal inflammatory disorder; malabsorption syndrome; gastric,
KW   sigmoid disease; antibacterial; antiviral; antiinflammatory;
KW   cytostatic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155302-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01240.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
```

```
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251998.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-465567/50.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the colon including colon cancers and also for
PT   testing and detection e.g. diagnosis -
XX
PS   Disclosure; SEQ ID No 518; 562pp; English.
XX
CC   The present invention relates to the isolation of novel human colon
CC   associated polypeptides (AAU22468-AAU22701), and the cDNA and genomic
CC   sequences encoding for them. The sequences of the invention are useful
CC   in the diagnosis, treatment, prevention and/or prognosis of disorders
CC   of the colon including colon cancer, congenital abnormalities
CC   (e.g. atresia and stenosis), bacterial and viral infections,
CC   inflammatory bowel disease (IBD), neoplastic cell disorders
CC   (e.g. polyps and adenomas, intestinal inflammatory disorders, colitis,
CC   colonic inflammation, diarrhoea and dysentery, malabsorption syndromes
CC   (e.g. lactose intolerance), intestinal obstruction and sigmoid diseases.
CC   The polynucleotides sequences of the invention can also be used in gene
CC   therapy. AAS39582-AAS40060 represent DNA sequences encoding for the
CC   novel human colon associated polypeptides of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 32152 BP; 9229 A; 7256 C; 7335 G; 8332 T; 0 other;

  Query Match             33.2%;  Score 165.8;  DB 22;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 2.5e-25;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251
                 ||||  |||  ||||  |||    ||  ||||||    ||||  |||||      ||
Db      28891 TGTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 28832

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
                ||  |||| ||   ||||  |||     |   |  |  ||  ||||   |  ||||   ||
Db      28831 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 28772

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
                ||  ||    ||  ||||||  ||||  ||     ||  |    ||||           || |
Db      28771 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 28712

Qy        372 TTTTTAAA-------------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
                 ||| |                    |||  ||  ||||  |||||||    |  |
Db      28711 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 28652

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
                |||  |||  |||  |||  ||||   ||     |    |   |||||   |||||  |||||
Db      28651 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 28594

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
                || ||  |||   |||  |||||  ||||
Db      28593 ACCTGGCCAACATGGTGAAACCCTGTC 28567


RESULT 14
AAK89020/c
ID   AAK89020 standard; DNA; 32152 BP.
XX
AC   AAK89020;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 2596.
XX
```

KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR* 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

```
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Disclosure; SEQ ID NO 2596; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention.
XX
SQ   Sequence 32152 BP; 9229 A; 7256 C; 7335 G; 8332 T; 0 other;

  Query Match             33.2%;  Score 165.8;  DB 22;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 2.5e-25;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251
Db      28891 TGTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 28832

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
Db      28831 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 28772

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
Db      28771 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 28712

Qy        372 TTTTTAAA-------------------TTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
Db      28711 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 28652

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
Db      28651 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 28594

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
Db      28593 ACCTGGCCAACATGGTGAAACCCTGTC 28567


RESULT 15
AAK91534
ID   AAK91534 standard; DNA; 32152 BP.
XX
AC   AAK91534;
XX
DT   05-NOV-2001  (first entry)
XX
DE   Human digestive system antigen genomic sequence SEQ ID NO: 5110.
XX
KW   Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW   ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW   digestive system disorder; Meckel's diverticulum; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155314-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01324.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
```

```
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-502630/55.
XX
PT   Polynucleotides encoding digestive system antigens, useful for
PT   diagnosing, treating, preventing and/or prognozing disorders of the
PT   digestive system, particularly cancer and cancer metastases -
XX
PS   Disclosure; SEQ ID NO 5110; 986pp; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human digestive system antigens. These can be used in the
CC   diagnosis, treatment and prevention of digestive system disorders,
CC   including cancer, Meckel's diverticulum, bacterial or parasitic
CC   infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC   ulcerative colitis. The present sequence is a genomic DNA fragment
CC   encoding a digestive system antigen of the invention.
XX
SQ   Sequence 32152 BP; 8332 A; 7335 C; 7256 G; 9229 T; 0 other;

  Query Match             33.2%;  Score 165.8;  DB 22;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 2.5e-25;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTATTTATTTATTTATTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251

Db       3262 TGTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 3321

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311

Db       3322 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 3381

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371

Db       3382 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 3441

Qy        372 TTTTTAAA-------------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412

Db       3442 TTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 3501

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472

Db       3502 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 3559

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499

Db       3560 ACCTGGCCAACATGGTGAAACCCTGTC 3586

Search completed: October 24, 2003, 19:47:13
Job time : 129.999 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

```
Run on:          October 24, 2003, 13:34:49 ; Search time 31.2916 Seconds
                                              (without alignments)
                                              7052.739 Million cell updates/sec

Title:           US-09-830-902-1_COPY_1_500
Perfect score:   500
Sequence:        1 taactgactctgctgtgttt..........aacatggtaaaaccatgtcg 500

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        569378 seqs, 220631566 residues

Total number of hits satisfying chosen parameters:      1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	167.2	33.4	2480	4	US-09-534-638-3	Sequence 3, Appli
	2	141.2	28.2	18000	4	US-09-657-346A-17	Sequence 17, Appl
	3	136.4	27.3	45716	4	US-08-965-048-5	Sequence 5, Appli
	4	136.4	27.3	45989	4	US-08-965-048-6	Sequence 6, Appli
	5	136.4	27.3	174493	4	US-09-804-471A-3	Sequence 3, Appli
c	6	133.6	26.7	1268	4	US-09-369-247-42	Sequence 42, Appl
c	7	133.6	26.7	75395	4	US-09-984-890-3	Sequence 3, Appli
c	8	132.6	26.5	319608	4	US-09-539-333D-1	Sequence 1, Appli
c	9	132.6	26.5	319608	4	US-09-679-409-1	Sequence 1, Appli
c	10	132.4	26.5	40000	4	US-09-780-049-18	Sequence 18, Appl
	11	132	26.4	74962	4	US-09-685-853A-3	Sequence 3, Appli
c	12	132	26.4	90541	4	US-09-759-359A-3	Sequence 3, Appli
c	13	131.6	26.3	7130	3	US-09-056-105-31	Sequence 31, Appl
	14	131.2	26.2	66804	4	US-09-740-041-3	Sequence 3, Appli
	15	131	26.2	1701	3	US-09-078-294-9	Sequence 9, Appli
	16	130.8	26.2	55298	4	US-09-491-356C-1	Sequence 1, Appli
	17	129.8	26.0	23187	4	US-09-499-522-1	Sequence 1, Appli
c	18	129.4	25.9	19650	4	US-09-819-989-3	Sequence 3, Appli
	19	129.4	25.9	80246	3	US-09-078-294-4	Sequence 4, Appli
	20	129.2	25.8	3885	1	US-08-688-145-1	Sequence 1, Appli
c	21	129.2	25.8	116592	4	US-09-818-512-3	Sequence 3, Appli
c	22	129	25.8	5375	3	US-08-757-223-7	Sequence 7, Appli
	23	129	25.8	80595	3	US-09-078-294-3	Sequence 3, Appli
	24	129	25.8	116592	4	US-09-818-512-3	Sequence 3, Appli
	25	128.4	25.7	3001	4	US-09-539-333D-218	Sequence 218, App
c	26	128	25.6	863	3	US-08-943-731-171	Sequence 171, App
c	27	128	25.6	17606	3	US-08-943-731-4	Sequence 4, Appli
	28	127.8	25.6	1764	4	US-09-620-312D-548	Sequence 548, App
	29	127.8	25.6	4704	2	US-08-476-062A-52	Sequence 52, Appl
	30	127.8	25.6	202001	4	US-09-734-674-3	Sequence 3, Appli
c	31	127.6	25.5	16063	4	US-09-801-052-3	Sequence 3, Appli
c	32	127	25.4	3460	4	US-09-904-615-44	Sequence 44, Appl
c	33	127	25.4	55298	4	US-09-491-356C-1	Sequence 1, Appli
c	34	126.8	25.4	62804	4	US-09-800-960-3	Sequence 3, Appli
c	35	126.4	25.3	896	3	US-08-943-731-31	Sequence 31, Appl
c	36	126.4	25.3	16609	3	US-08-943-731-1	Sequence 1, Appli
c	37	126.4	25.3	99500	4	US-09-796-096-10	Sequence 10, Appl
	38	126.2	25.2	148567	4	US-09-801-876B-3	Sequence 3, Appli
	39	126	25.2	50000	4	US-09-146-053-4	Sequence 4, Appli
	40	126	25.2	319608	4	US-09-539-333D-1	Sequence 1, Appli
	41	126	25.2	319608	4	US-09-679-409-1	Sequence 1, Appli
c	42	125.6	25.1	55827	4	US-09-813-133A-3	Sequence 3, Appli
	43	125.4	25.1	1000	3	US-09-018-584A-33	Sequence 33, Appl
	44	125.4	25.1	2099	3	US-08-938-669A-5	Sequence 5, Appli
	45	125.4	25.1	2099	4	US-09-306-828-5	Sequence 5, Appli

ALIGNMENTS

```
RESULT 1
US-09-534-638-3/c
; Sequence 3, Application US/09534638
; Patent No. 6320038
; GENERAL INFORMATION:
;  APPLICANT: Panula, Pertti A.J.
;  APPLICANT: Brandt, Annika
;  APPLICANT: Westerlund, Johanna
;  TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
;  TITLE OF INVENTION: for therapy and diagnosis
;  FILE REFERENCE: 2530-104
;  CURRENT APPLICATION NUMBER: US/09/534,638
;  CURRENT FILING DATE: 2000-03-27
;  EARLIER APPLICATION NUMBER: 09/365755
;  EARLIER FILING DATE: 1999-08-03
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
;   LENGTH: 2480
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-534-638-3

  Query Match             33.4%;  Score 167.2;  DB 4;  Length 2480;
  Best Local Similarity   75.6%;  Pred. No. 2.7e-30;
  Matches  239;  Conservative    0;  Mismatches   63;  Indels   14;  Gaps    2;

Qy        193 ATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTC-TGTTGCCCAGGCTGGAATGCA 251

Db       1885 AATTTAACTTTTTTTCTTTTTGAGATGGAGTTTCGCTCTTTTGCCCAGGCTGGAGTGCA 1826

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311

Db       1825 GTGGCGCGATCTCGGCTCACTGCAACCTCTGGCTCCTGGGTTCAAGTGATTCTCCTGCCT 1766

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371

Db       1765 CAGCCTCCCAAGTAGCTGGGACTACAGGCGTCAACCACCATGCCCAGCCTGTATCTTTTT 1706

Qy        372 TTTTTAAA-------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCA 418

Db       1705 TTTTTAATTAAAAACCATTTTTTTTTTGCCGGGTGCAGTGCCTCACACCTGTAACCCCA 1646

Qy        419 GCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGG 478

Db       1645 GCACTTTGGGAGACCAAGGCGGGTGGATCACCTGAGGCAGGAGTTCGAGACCAGCCTGG 1586

Qy        479 CCAACATGGTAAAACC 494

Db       1585 CCAACATGGTGAAACC 1570
```

```
RESULT 2
US-09-657-346A-17
; Sequence 17, Application US/09657346A
; Patent No. 6503754
; GENERAL INFORMATION:
;  APPLICANT: Hong Zhang
;  APPLICANT:  Jacqueline Wyatt
;  TITLE OF INVENTION: ANTISENSE MODULATION OF BH3 INTERACTING DOMAIN DEATH AGONIST
;  TITLE OF INVENTION:  EXPRESSION
;  FILE REFERENCE: RTS-0135
;  CURRENT APPLICATION NUMBER: US/09/657,346A
;  CURRENT FILING DATE:  2000-09-07
;  NUMBER OF SEQ ID NOS: 174
; SEQ ID NO 17
;   LENGTH: 18000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2144)...(2155)
;   NAME/KEY: CDS
;   LOCATION: (8247)...(8457)
;   NAME/KEY: CDS
;   LOCATION: (12772)...(12911)
;   NAME/KEY: CDS
;   LOCATION: (14031)...(14243)
;   NAME/KEY: CDS
;   LOCATION: (16669)...(16680)
US-09-657-346A-17

  Query Match             28.2%;  Score 141.2;  DB 4;  Length 18000;
  Best Local Similarity   74.7%;  Pred. No. 4.5e-24;
  Matches  218;  Conservative    0;  Mismatches   68;  Indels    6;  Gaps    3;

Qy        203 ATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATC 262
Db      11355 ACTTTTTTTAAGAAACGGGGTCTTGCTCTGTTGCCCAGGCTGGAGTGTAGTGGTATGATC 11414

Qy        263 TCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAA 322
Db      11415 ACAGCTCACTGCAGCCTTGAATTCCTGGCCTCAAGGAGTCCTCCCACCTCAGCCTCCTGT 11474

Qy        323 GTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTT 382
Db      11475 GTAGCTGGGATTATAGGTACGTGC-GGTACACCTGGCTCCCTTTAAAAGTTTTT--TGT 11530

Qy        383 TGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGC 442
Db      11531 AGAGGCAGGGCACAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCAAGGCAGGA 11590

Qy        443 CGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACC 494
Db      11591 GGATCAC--AAGGTCAGGAGTTCGAGACCAGCCTGACCAACATGGTGAAACC 11640

RESULT 3
US-08-965-048-5
; Sequence 5, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
;  APPLICANT: Chen, Hong
;  APPLICANT: Freimer, Nelson
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
;  FILE REFERENCE: 7853-093
;  CURRENT APPLICATION NUMBER: US/08/965,048
;  CURRENT FILING DATE: 1997-11-05
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 45716
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-965-048-5

  Query Match             27.3%;  Score 136.4;  DB 4;  Length 45716;
  Best Local Similarity   71.2%;  Pred. No. 7.2e-23;
  Matches  195;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    1;

Qy        106 CATTTTCACTGAAGTAACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGA 165
Db      34614 CAATTTGAAGGTTAATAGACATGGTTCCTATGCTTTGTAAATACAGAAATGTGACAGCAT 34673

Qy        166 GTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCT---AGATGGAG 222
Db      34674 TTTTTTTTTTTGTTTTTTGGTGGTTTTTTTTGTTTTGTTTTGTTTTGTTTTGAGATAGAG 34733

Qy        223 TCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCA 282
Db      34734 TCTTACTCTGTTGCCCAGGCTGGAGTGCAATGGCATGATCTCGGCTCACTGCAACCTCTG 34793

Qy        283 CCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAAT 342
Db      34794 CCTCCCGAGTTCAAGCAATTCTTCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGATGT 34853

Qy        343 GAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTT 376
Db      34854 GTGCCACCATGCCTGGCTTTTTTTTTTTTTTTTT 34887

RESULT 4
US-08-965-048-6
; Sequence 6, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
;  APPLICANT: Chen, Hong
;  APPLICANT: Freimer, Nelson
;  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
;  TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
;  FILE REFERENCE: 7853-093
;  CURRENT APPLICATION NUMBER: US/08/965,048
;  CURRENT FILING DATE: 1997-11-05
;  NUMBER OF SEQ ID NOS: 8
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
;   LENGTH: 45989
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-08-965-048-6

  Query Match             27.3%;  Score 136.4;  DB 4;  Length 45989;
  Best Local Similarity   71.2%;  Pred. No. 7.2e-23;
  Matches  195;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    1;

Qy        106 CATTTTCACTGAAGTAACAAATCCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGA 165
Db      34728 CAATTTGAAGGTTAATAGACATGGTTCCTATGCTTTGTAAATACAGAAATGTGACAGCAT 34787

Qy        166 GTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCT---AGATGGAG 222
Db      34788 TTTTTTTTTTTGTTTTTTGGTGGTTTTTTTTGTTTTGTTTTGTTTTGTTTTGAGATAGAG 34847

Qy        223 TCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCA 282
Db      34848 TCTTACTCTGTTGCCCAGGCTGGAGTGCAATGGCATGATCTCGGCTCACTGCAACCTCTG 34907

Qy        283 CCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAAT 342
Db      34908 CCTCCCGAGTTCAAGCAATTCTTCTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGATGT 34967

Qy        343 GAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTT 376
Db      34968 GTGCCACCATGCCTGGCTTTTTTTTTTTTTTTTT 35001
```

```
RESULT 5
US-09-804-471A-3
; Sequence 3, Application US/09804471A
; Patent No. 6479269
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001164
;  CURRENT APPLICATION NUMBER: US/09/804,471A
;  CURRENT FILING DATE:  2001-03-13
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 174493
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(174493)
;   OTHER INFORMATION: n = A,T,C or G
US-09-804-471A-3

  Query Match             27.3%;  Score 136.4;  DB 4;  Length 174493;
  Best Local Similarity   68.1%;  Pred. No. 9.4e-23;
  Matches  207;  Conservative    0;  Mismatches   91;  Indels    6;  Gaps    1;

Qy        195 TTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATG 254

Db      91427 TATTTGTTTTTTTTTTAAGAGACAAGGTCTTGCTGTGTTACCCAGACTGGAGTGCAGTG 91486

Qy        255 GCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAA 314

Db      91487 GTACAATCGTAGCTCACTGTAGCCTTGAACTCCTGGGCTCGAGCGATCCTCCCACCCCAG 91546

Qy        315 TCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTT 374

Db      91547 CCTCCCACGTAGCTGGGACTAGAGGCATGAGCCACTATGCCCGACTCATTTTTAAACATT 91606

Qy        375 TTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCA 434

Db      91607 TTTTATAGAGACG------GCTGGGTGTGGTACCTGTAATCCCAGCACTTTGGGAGACTG 91660

Qy        435 AGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACC 494

Db      91661 AGGCGGGCGGATTGCTTGAGCCCAGGAATTTGAGACCAGCCTGGGCAACATGGAGAAACC 91720

Qy        495 ATGT 498

Db      91721 CCGT 91724


RESULT 6
US-09-369-247-42/c
; Sequence 42, Application US/09369247
; Patent No. 6569992
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: 44 Human Secreted Proteins
;  FILE REFERENCE: PZ024P1
;  CURRENT APPLICATION NUMBER: US/09/369,247
;  CURRENT FILING DATE: 1999-08-05
;  EARLIER APPLICATION NUMBER: 60/074,118
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/074,157
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/074,137
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/074,341
;  EARLIER FILING DATE: 1998-02-09
;  EARLIER APPLICATION NUMBER: 60/074,141
;  EARLIER FILING DATE: 1998-02-09
;  NUMBER OF SEQ ID NOS: 172
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
;   LENGTH: 1268
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (2)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (15)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (23)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (368)
;   OTHER INFORMATION: n equals a,t,g, or c
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (479)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-369-247-42

  Query Match             26.7%;  Score 133.6;  DB 4;  Length 1268;
  Best Local Similarity   78.4%;  Pred. No. 1.6e-22;
  Matches  160;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232

Db       1260 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTTGCTCTG 1201

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292

Db       1200 TCGCCCAGGCTGGAGTGCAGTGGCGCGATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGC 1141

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352

Db       1140 TCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCGCCCGCTACCAC 1081

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTT 376

Db       1080 GCCCGGCTAATTTTTTGTATTTTT 1057


RESULT 7
US-09-984-890-3/c
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
;  APPLICANT: YAN, Chunhua et al.
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001306
;  CURRENT APPLICATION NUMBER: US/09/984,890
;  CURRENT FILING DATE:  2001-10-31
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 75395
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(75395)
;   OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3
```

```
  Query Match             26.7%;  Score 133.6;  DB 4;  Length 75395;
  Best Local Similarity   72.6%;  Pred. No. 3.5e-22;
  Matches  215;  Conservative    0;  Mismatches   74;  Indels    7;  Gaps    3;

Qy        204 TTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCT 263

Db      10020 TATTTATCTTGAGATGGGGTCTCACTCTGTTGCCCAGGCTGGAATGCAGTGGCATGATCA 9961

Qy        264 CAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAG 323

Db       9960 CAGCTCACTGCAGCCTCAACTTCCCAGGCTCAAGTGATCCTCACATCTCAGCCTTCTGAG 9901

Qy        324 TAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTT 383

Db       9900 TAGGTGGGA-TACACGCATGTGCTACCACAC----CCAGGTAACTTTTTAAAAATTTTTT 9846

Qy        384 GAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCC 443

Db       9845 TGGGCTGGGCGCGGTGGCTCCCGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGGTG 9786

Qy        444 GATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499

Db       9785 GATCAC--GAGGTCAGAAGATCGAGACCATCCTGGCTAACAAGGTGAAACCCCGTC 9732
```

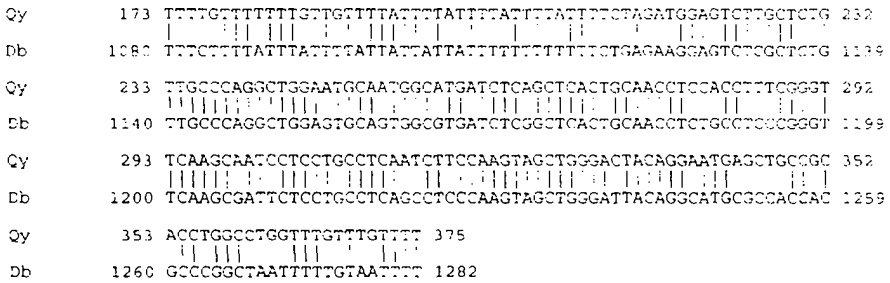
RESULT 8
US-09-539-333D-1/c
; Sequence 1, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 31..1107
;   OTHER INFORMATION: 5'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 1108..1289
;   OTHER INFORMATION: exon A g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 14877..14920
;   OTHER INFORMATION: exon B g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 18778..18862
;   OTHER INFORMATION: exon Bbis g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 25593..25740
;   OTHER INFORMATION: exon C g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29388..29502
;   OTHER INFORMATION: exon D g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 29967..30282
;   OTHER INFORMATION: exon E g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 64666..64812
;   OTHER INFORMATION: exon F g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 65505..65853
;   OTHER INFORMATION: exon G g35018 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 65854..67854
;   OTHER INFORMATION: 3'regulatory region g35018 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 94124..94964
;   OTHER INFORMATION: exon g35017
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 201188..201234
;   OTHER INFORMATION: exon S g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U g35030 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216836..216915
;   OTHER INFORMATION: exon V g35030 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 213818..215818
;   OTHER INFORMATION: 3'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215941
;   OTHER INFORMATION: exon R complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 215819..215975
;   OTHER INFORMATION: exon Rbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..216952
;   OTHER INFORMATION: exon Qbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 216661..217061
;   OTHER INFORMATION: exon Q complement g34872 gene
;   FEATURE:

```
;   NAME/KEY: exon
;   LOCATION: 217027..217061
;   OTHER INFORMATION: exon Q1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 229647..229742
;   OTHER INFORMATION: exon X complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 230408..230721
;   OTHER INFORMATION: exon P complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231272..231412
;   OTHER INFORMATION: exon Obis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231787..231880
;   OTHER INFORMATION: exon O2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 231870..231879
;   OTHER INFORMATION: exon O1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 234174..234321
;   OTHER INFORMATION: exon O complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 237406..237428
;   OTHER INFORMATION: exon Nbis complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239807
;   OTHER INFORMATION: exon N2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 239719..239853
;   OTHER INFORMATION: exon N complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240569
;   OTHER INFORMATION: exon M1117 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240596
;   OTHER INFORMATION: exon M1090 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240617
;   OTHER INFORMATION: exon M1069 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240644
;   OTHER INFORMATION: exon MS2 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240824
;   OTHER INFORMATION: exon M862 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..240994
;   OTHER INFORMATION: exon M692 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240528..241685
;   OTHER INFORMATION: exon M1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 240800..240993
;   OTHER INFORMATION: exon MS1 complement g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 241686..243685
;   OTHER INFORMATION: 5'regulatory region g34872 gene
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 290652..292652
;   OTHER INFORMATION: 3'regulatory region g34665 gene
;   FEATURE:
;   NAME/KEY: exon
;   LOCATION: 292653..292841

  Query Match             26.5%;  Score 132.6;  DB 4;  Length 319608;
  Best Local Similarity   76.8%;  Pred. No. 8.1e-22;
  Matches  162;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        173 TTTTGTTTTTTGTTGTTTTATTTTATTTTATTTTATTTCTAGATGGAGTCTTGCTCTG 232
              ||||  ||| ||     |  | ||   | |     | |    |||||   ||||    |||
Db      16340 TTTTCGTTTGTTTGTTTGTTTTTGTTTTGTTTTGTTTTTTTTAGATGGAGTCTTGCTCTG 16281

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
              |||| ||| |||    || | |||  |||||  ||||   |||||  |    ||    ||||
Db      16280 TTGCTCAGGCTGGAGTGTAGTGGCGTGATCTCAGCTCACTGCAACCTCTGCCTCCCGGGT 16221

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
               ||||  |||     ||  |||     | ||  | ||||||    | ||||          |  
Db      16220 TCAAGCAATTCTCATGACTCAGCCTCCCAAGTAGCTGGGAATACAGGCGCATGCCACCAT 16161

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTAAATTTT 383
               |||||      ||     || ||  | | |
Db      16160 GCCTGGCTAATTTTTTTTTTTTTGTATTTT 16130


RESULT 9
US-09-679-409-1/c
; Sequence 1, Application US/09679409
; Patent No. 6555316
; GENERAL INFORMATION:
;  APPLICANT:  Cohen, Daniel
;  APPLICANT:  Blumenfeld, Marta
;  APPLICANT:  Chumakov, Ilya
;  APPLICANT:  Bougueleret, Lydie
;  APPLICANT:  Essioux, Laurent
;  TITLE OF INVENTION:  SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
;  FILE REFERENCE:  53.US15.CIP
;  CURRENT APPLICATION NUMBER: US/09/679,409
;  CURRENT FILING DATE:  2000-10-03
;  PRIOR APPLICATION NUMBER:  09/539,333
;  PRIOR FILING DATE:  2000-03-03
;  PRIOR APPLICATION NUMBER:  09/416,384
;  PRIOR FILING DATE:  1999-10-12
;  PRIOR APPLICATION NUMBER:  60/168,088
;  PRIOR FILING DATE:  1999-11-30
;  NUMBER OF SEQ ID NOS:  134
;  SOFTWARE:  Patent.pm
; SEQ ID NO 1
;   LENGTH: 319608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 199122..201122
;   OTHER INFORMATION: 5'regulatory region
;   NAME/KEY: exon
;   LOCATION: 201123..201234
;   OTHER INFORMATION: exon S
;   NAME/KEY: exon
;   LOCATION: 201123..201560
;   OTHER INFORMATION: exon S2
;   NAME/KEY: exon
;   LOCATION: 214676..214793
;   OTHER INFORMATION: exon T
;   NAME/KEY: exon
;   LOCATION: 215702..215746
;   OTHER INFORMATION: exon U
```

; NAME/KEY: exon
; LOCATION: 216836..216994
; OTHER INFORMATION: exon V
; NAME/KEY: exon
; LOCATION: 216836..217077
; OTHER INFORMATION: exon V2
; NAME/KEY: exon
; LOCATION: 217671..217764
; OTHER INFORMATION: exon V1
; NAME/KEY: exon
; LOCATION: 227655..227736
; OTHER INFORMATION: exon V4
; NAME/KEY: exon
; LOCATION: 238715..238919
; OTHER INFORMATION: exon V3
; NAME/KEY: exon
; LOCATION: 240440..240673
; OTHER INFORMATION: exon W
; NAME/KEY: exon
; LOCATION: 240440..241153
; OTHER INFORMATION: exon W2
; NAME/KEY: exon
; LOCATION: 241072..241291
; OTHER INFORMATION: exon X
; NAME/KEY: exon
; LOCATION: 244353..244561
; OTHER INFORMATION: exon Y
; NAME/KEY: exon
; LOCATION: 246273..247802
; OTHER INFORMATION: exon Z
; NAME/KEY: misc_feature
; LOCATION: 247803..249803
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 8316
; OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 21672
; OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 65485
; OTHER INFORMATION: 8-128-33 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 95396
; OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 107281
; OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 160640
; OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 162876
; OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 168974
; OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 170810
; OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 173358
; OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 189957
; OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 197163
; OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 200778
; OTHER INFORMATION: 8-303-235 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202651
; OTHER INFORMATION: 8-300-221 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 202679
; OTHER INFORMATION: 8-300-193 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 203378
; OTHER INFORMATION: 8-299-128 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204138
; OTHER INFORMATION: 8-296-213 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 204605
; OTHER INFORMATION: 8-252-190 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 204934
; OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 205206
; OTHER INFORMATION: 8-295-248 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 205329
; OTHER INFORMATION: 8-295-125 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 206064
; OTHER INFORMATION: 8-293-130 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 206545
; OTHER INFORMATION: 8-292-198 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 207313
; OTHER INFORMATION: 8-251-322 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 208285
; OTHER INFORMATION: 8-289-322 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 208960
; OTHER INFORMATION: 8-287-249 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 209123
; OTHER INFORMATION: 8-287-86 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 209631
; OTHER INFORMATION: 8-285-319 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210361
; OTHER INFORMATION: 8-283-278 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210463
; OTHER INFORMATION: 8-283-176 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 210486
; OTHER INFORMATION: 8-283-153 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210583
; OTHER INFORMATION: 8-283-56 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 210879
; OTHER INFORMATION: 8-282-345 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 210964
; OTHER INFORMATION: 8-282-260 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 210979
; OTHER INFORMATION: 8-282-245 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 211050
; OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGGAAGA
; NAME/KEY: allele
; LOCATION: 211132
; OTHER INFORMATION: 8-282-92 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 211247

```
;   OTHER INFORMATION: 8-281-367  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 211315
;   OTHER INFORMATION: 8-281-299  : polymorphic base A or G
;   NAME/KEY: allele
;   LOCATION: 211366
;   OTHER INFORMATION: 8-281-248  : polymorphic base G or C
;   NAME/KEY: allele
;   LOCATION: 212520
;   OTHER INFORMATION: 8-279-197  : polymorphic base A or C
;   NAME/KEY: allele
;   LOCATION: 212821
;   OTHER INFORMATION: 8-278-289  : polymorphic base C or T
;   NAME/KEY: allele

  Query Match           26.5%;  Score 132.6;  DB 4;  Length 319608;
  Best Local Similarity   76.8%;  Pred. No. 8.1e-22;
  Matches  162;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232
Db      16340 TTTTCGTTTGTTTGTTTGTTTTTGTTTTGTTTTGTTTTTTTTAGATGGAGTCTTGCTCTG 16281

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
Db      16280 TTGCTCAGGCTGGAGTGTAGTGGCGTGATCTCAGCTCACTGCAACCTCTGCCTCCCGGT 16221

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
Db      16220 TCAAGCAATTCTCATGACTCAGCCTCCCAAGTAGCTGGGAATACAGGCGCATGCCACCAT 16161

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTTAAATTTT 383
Db      16160 GCCTGGCTAATTTTTTTTTTTTTGTATTTT 16130


RESULT 10
US-09-780-049-18/c
; Sequence 18, Application US/09780049
; Patent No. 6465250
; GENERAL INFORMATION:
;  APPLICANT:  Brett P. Monia
;  APPLICANT:  Jacqueline Wyatt
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF PROTEIN PHOSPHATASE 2 CATALYTIC SUBUNIT
;  TITLE OF INVENTION:  EXPRESSION
;  FILE REFERENCE:  RTS-0134
;  CURRENT APPLICATION NUMBER: US/09/780,049
;  CURRENT FILING DATE:  2001-02-09
;  NUMBER OF SEQ ID NOS:  96
; SEQ ID NO 18
;   LENGTH: 40000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-780-049-18

  Query Match           26.5%;  Score 132.4;  DB 4;  Length 40000;
  Best Local Similarity   69.3%;  Pred. No. 5.9e-22;
  Matches  223;  Conservative    0;  Mismatches   96;  Indels    3;  Gaps    3;

Qy        180 TTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCA 239
Db      12464 TGTCTAATATTTATTTACTTTTTTTTTTTTTTTGAGACAAGTTCTTGCTCTGTTCCCCA 12405

Qy        240 GGC-TGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGC 298
Db      12404 GGCTTCGAGTGCTGTAGCATGATCACAGCTCATTGCAGTCTAGATCTCCTGGACCCAAGC 12345

Qy        299 AATCCTCCTGCCTCAATCTTCCA-AGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTG 357
Db      12344 AATCCTCTTGCCTCAGCCTCACAGAGTAGCTAGAACCACAGGTGTG-CCACCACACCCAA 12286

Qy        358 GCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCC 417
Db      12285 CTAATTTATTTTTAAAAATTATTCGTAGGGCTGGGTGCAGTGGCTCACGCCTGTAATCCC 12226

Qy        418 AGCACTTTGGGAGACCAAGGCAGGCGGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAG 477
Db      12225 AGCACTTCGGGAGGCCGAGGCAGGTGGATCACTTGAGGTCAGGAGTTCAAGACCAGCCTG 12166

Qy        478 GCCAACATGGTAAAACCATGTC 499
Db      12165 GCCAAAGTGGTGAAACCCCATC 12144


RESULT 11
US-09-685-853A-3
; Sequence 3, Application US/09685853A
; Patent No. 6479270
; GENERAL INFORMATION:
;  APPLICANT: WEI, Ming-Hui et al.
;  TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL000871
;  CURRENT APPLICATION NUMBER: US/09/685,853A
;  CURRENT FILING DATE:  2000-05-06
;  PRIOR APPLICATION NUMBER: 60/182,194
;  PRIOR FILING DATE: 2000-02-14
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 74962
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(74962)
;   OTHER INFORMATION: n = A,T,C or G
US-09-685-853A-3

  Query Match           26.4%;  Score 132;  DB 4;  Length 74962;
  Best Local Similarity   73.7%;  Pred. No. 8.3e-22;
  Matches  168;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy        166 GTTTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCT 225
Db      42109 GTTTTTAATTTTGTTTTTTTTTGTTGTTGTTGTTTTTTTTTTTTTTTTGAGATGGAGTCT 42168

Qy        226 TGCTCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCT 285
Db      42169 CACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACCTCTGCCT 42228

Qy        286 TTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAG 345
Db      42229 CCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCATGTG 42288

Qy        346 CTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGT 393
Db      42289 CCACCACACCCAGCTAATTTTTTTTTTTTTTTTTTTTTGGACAGAGT 42336


RESULT 12
US-09-759-359A-3/c
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:
;  APPLICANT: ABU-THREIDEH, Jane et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001043
;  CURRENT APPLICATION NUMBER: US/09/759,359A
;  CURRENT FILING DATE:  2001-01-16
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 3
;   LENGTH: 90541
;   TYPE: DNA
;   ORGANISM: Human
US-09-759-359A-3

  Query Match            26.4%;  Score 132;  DB 4;  Length 90541;
  Best Local Similarity  67.8%;  Pred. No. 8.7e-22;
  Matches  236;  Conservative    0;  Mismatches   95;  Indels   17;  Gaps    3;

Qy        168 TTTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTG 227
Db      54297 TTGAATTATGATTTTTCATCACTTTTAAAAATTAATTTCTTTTTAGAGATGGGATCCCG 54238

Qy        228 CTCTGTTGCCCAGGCTGGAATGCAATGGCA-----------TGATCTCAGCTCACTGCAA 276
Db      54237 CTATGTT-CCCAGGCTGGAGTGCAGTGGCTATTCAGAGGTGTAAGTAAAGCACACTGAAG 54179

Qy        277 CCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTAC 336
Db      54178 CCTCAAGCTTCTGGCCTCAAGTGATCCTCTTGCTTCAGCCTCCCAAGTGGCTGGGATTAC 54119

Qy        337 AGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTAAATTT-----TGAGGCCAG 391
Db      54118 AGGTGTGAGCCACCACATCCAGCTTCCTAAGTCACTTATTAAAGTTCTCTGGGTTGGCTG 54059

Qy        392 GTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTT 451
Db      54058 GCACGATGGCTCACACCTGTAATCCCAGCACTTTGGGAGGTCAAGGCAGGTGGATCACCT 53999

Qy        452 GAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
Db      53998 GAGGTCTGGAGTTCAAGACCAGCCTGGCCAACACGGTGAAACCCCCTC 53951

RESULT 13
US-09-056-105-31/c
; Sequence 31, Application US/09056105
; Patent No. 6287569
; GENERAL INFORMATION:
;  APPLICANT: KIPPS, THOMAS J.
;  APPLICANT: WU, YUNQI
;  TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR
;  TITLE OF INVENTION: PROCESSING
;  FILE REFERENCE: 233/221
;  CURRENT APPLICATION NUMBER: US/09/056,105
;  CURRENT FILING DATE: 1998-04-06
;  EARLIER APPLICATION NUMBER: 60/043,467
;  EARLIER FILING DATE: 1997-04-10
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 31
;   LENGTH: 7130
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-056-105-31

  Query Match            26.3%;  Score 131.6;  DB 3;  Length 7130;
  Best Local Similarity  70.4%;  Pred. No. 6.5e-22;
  Matches  176;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy        128 CCTCCATAAACTAGAAAATCTCAAACTGGTGACTGGGAGTTTTGGTTTTGTTTTTTTGTT 187
Db       2967 CCCTCATACCCACACACATCAGGGATAAAAGATTCCTTTTCTTTTCTTTTCTTTTCTTTT 2908

Qy        188 GTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAA 247
Db       2907 TTTTTTTTTTTTTTTTTTTTTTTTTGAGGCGGAGTCTCACTCTGTCACCCAGGCTGGAG 2848

Qy        248 TGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCT 307
Db       2847 TGCAGTGGCACGGTCTCGGCTCACTGCAACCTCAACCTCCCAGGTTCAAGCAATTCTCCT 2788

Qy        308 GCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTG 367
Db       2787 CCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCATGCGCCACCACGCCTGGCTAATTTTT 2728

Qy        368 TTTGTTTTTA 377
Db       2727 TGTATTTTTA 2718

RESULT 14
US-09-740-041-3
; Sequence 3, Application US/09740041
; Patent No. 6562593
; GENERAL INFORMATION:
;  APPLICANT: MERKULOV, Karl et al
;  TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;  TITLE OF INVENTION:  AND USES THEREOF
;  FILE REFERENCE: CL001001
;  CURRENT APPLICATION NUMBER: US/09/740,041
;  CURRENT FILING DATE: 2000-12-20
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 66804
;   TYPE: DNA
;   ORGANISM: Human
US-09-740-041-3

  Query Match            26.2%;  Score 131.2;  DB 4;  Length 66804;
  Best Local Similarity  76.9%;  Pred. No. 1.3e-21;
  Matches  160;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy        173 TTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTG 232
Db      52221 TATATTTTTTTTGTTTGTTTGTTTTGTTTTGTTTTGTTTTTGAGATGGAGTCTCGCTCTG 52280

Qy        233 TTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGT 292
Db      52281 TCACCCAGGCTGGAGTGCAGTGGCGTGATCTCAGCTCACTGCAAGTTCCGCCTCCCAGGT 52340

Qy        293 TCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGC 352
Db      52341 TCACACCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGCACCCGCCACCAC 52400

Qy        353 ACCTGGCCTGGTTTGTTTGTTTTAAAT 380
Db      52401 ACCCGGCTAATTTTTTGTATTTTTTAGT 52428

RESULT 15
US-09-078-294-9
; Sequence 9, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
;  APPLICANT: Choo, Kong-Hong Andy
;  APPLICANT: Du Sart, Desiree
;  APPLICANT: Cancilla, Michael R.
;  TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;  FILE REFERENCE: Davies Col
;  CURRENT APPLICATION NUMBER: US/09/078,294
;  CURRENT FILING DATE: 1998-05-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
;   LENGTH: 1701
;   TYPE: DNA
;   ORGANISM: BAC-F2 contig 5
US-09-078-294-9

  Query Match            26.2%;  Score 131;  DB 3;  Length 1701;
  Best Local Similarity  77.6%;  Pred. No. 6.7e-22;
  Matches  158;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;
```



Search completed: October 24, 2003, 15:42:43
Job time : 36.2916 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 123.169 Seconds
(without alignments)
10886.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: 1 taactgacttctgctgtgttt.....aacatggtaaaaccatgtcg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	180.4	36.1	877	13	US-10-027-632-139570	Sequence 139570,
2	180.4	36.1	877	13	US-10-027-632-139571	Sequence 139571,
3	179.2	35.8	624	13	US-10-027-632-139569	Sequence 139569,
4	175.6	35.1	39725	12	US-10-017-161-1611	Sequence 1611, Ap
c 5	169.8	34.0	871	12	US-10-017-161-1849	Sequence 1849, Ap
c 6	166.6	33.3	36400	12	US-10-225-810-10	Sequence 10, Appl
c 7	166.6	33.3	250000	12	US-10-225-810-26	Sequence 26, Appl
8	165.8	33.2	32152	10	US-09-764-855-328	Sequence 328, App
c 9	165.8	33.2	32152	11	US-09-764-872-518	Sequence 518, App
10	165.8	33.2	32152	14	US-10-072-349-328	Sequence 328, App
11	165.8	33.2	32204	10	US-09-764-855-327	Sequence 327, App
c 12	165.8	33.2	32204	11	US-09-764-872-517	Sequence 517, App
13	165.8	33.2	32204	14	US-10-072-349-327	Sequence 327, App
14	162.2	32.4	662	13	US-10-027-632-276723	Sequence 276723,
c 15	162.2	32.4	798	13	US-10-027-632-147363	Sequence 147363,
c 16	162.2	32.4	13024	11	US-09-764-891-7427	Sequence 7427, Ap
c 17	162.2	32.4	46130	12	US-10-017-161-985	Sequence 985, App
18	162.2	32.4	55611	12	US-10-017-161-783	Sequence 783, App
19	161	32.2	35425	12	US-10-017-161-2429	Sequence 2429, Ap
c 20	160.4	32.1	3459	9	US-09-764-887-412	Sequence 412, App
c 21	160.4	32.1	3459	14	US-10-073-961-412	Sequence 412, App
c 22	160.2	32.0	601	13	US-10-027-632-189758	Sequence 189758,
c 23	160.2	32.0	601	13	US-10-027-632-189759	Sequence 189759,
c 24	159.8	32.0	25001	12	US-10-017-161-2063	Sequence 2063, Ap
c 25	157.8	31.6	2442	13	US-10-027-632-110068	Sequence 110068,
c 26	157.8	31.6	2442	13	US-10-027-632-110069	Sequence 110069,
27	154.4	30.9	650	13	US-10-027-632-236200	Sequence 236200,
c 28	153.8	30.8	20951	12	US-09-805-455-3	Sequence 3, Appli
c 29	153.6	30.7	10855	11	US-09-764-891-7965	Sequence 7965, Ap
c 30	153.6	30.7	15714	9	US-09-764-869-2355	Sequence 2355, Ap
c 31	153.6	30.7	15714	14	US-10-091-504-2355	Sequence 2355, Ap
c 32	153.2	30.6	32194	11	US-09-764-891-7028	Sequence 7028, Ap
33	152.4	30.5	10236	9	US-09-764-853-878	Sequence 878, App
c 34	152.4	30.5	10236	14	US-10-091-438-242	Sequence 242, App
c 35	152.4	30.5	10236	14	US-10-091-438-282	Sequence 282, App
c 36	151.6	30.3	2280	13	US-10-027-632-103591	Sequence 103591,
c 37	151.6	30.3	2280	13	US-10-027-632-103592	Sequence 103592,
c 38	151.4	30.3	21129	9	US-09-764-669-1734	Sequence 1734, Ap
c 39	151.4	30.3	21129	14	US-10-091-504-1734	Sequence 1734, Ap
40	150.8	30.2	63720	14	US-10-105-637-4	Sequence 4, Appli
c 41	150.4	30.1	110096	10	US-09-880-107-1542	Sequence 1542, Ap
c 42	149.8	30.0	1282	10	US-09-822-830A-319	Sequence 319, App
43	149.2	29.8	53542	9	US-09-801-574-61	Sequence 61, Appl
44	148.6	29.7	770	13	US-10-027-632-170830	Sequence 170830,
45	148.6	29.7	770	13	US-10-027-632-170831	Sequence 170831,

ALIGNMENTS

RESULT 1
US-10-027-632-139570
; Sequence 139570, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,302
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 139570
; LENGTH: 877
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-139570

Query Match 36.1%; Score 180.4; DB 13; Length 877;
Best Local Similarity 73.8%; Pred. No. 1.6e-36;
Matches 242; Conservative 1; Mismatches 82; Indels 3; Gaps 1;

Qy 175 TTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTT 234
Db 234 TTTTTTGTGTGTTTTTTGTTTTGTTTTGCTTTTGTTTTTGAGACAGGGTCTCACTCTGCT 293

```
Qy        235 GCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTC 294
Db        294 GCCCAGGCTGAAGTGCAATGGGGTGATCATGGCTCACTGCAGTCTCAACCTTCTGGGCTC 353

Qy        295 AAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCAC 354
Db        354 AAGCAATCCTCCCACCTCAGCTTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCATGC 413

Qy        355 CTGGCCTGGTTTGTTTGTTTTTAA---ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGT 411
Db        414 ACGGCTGTAAAAAGTACCTTTTAAAGGAGAATAAAGCCAGGAGCAGTGGCTCACRTCTGT 473

Qy        412 GATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACC 471
Db        474 AATCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACC 533

Qy        472 AGCCAGGCCAACATGGTAAAACCATGTC 499
Db        534 AGCCTGGCCAACATGGCAAAACCCCATC 561


RESULT 2
US-10-027-632-139571
; Sequence 139571, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 139571
;   LENGTH: 877
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-139571

  Query Match           36.1%;  Score 180.4;  DB 13;  Length 877;
  Best Local Similarity   73.8%;  Pred. No. 1.6e-36;
  Matches  242;  Conservative    1;  Mismatches   82;  Indels    3;  Gaps    1;

Qy        175 TTGTTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTT 234
Db        234 TTTTTTGTGTGTTTTTTGTTTTGTTTTGCTTTTGTTTTTGAGACAGGGTCTCACTCTGCT 293

Qy        235 GCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTC 294
Db        294 GCCCAGGCTGAAGTGCAATGGGGTGATCATGGCTCACTGCAGTCTCAACCTTCTGGGCTC 353

Qy        295 AAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCAC 354
Db        354 AAGCAATCCTCCCACCTCAGCTTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCATGC 413

Qy        355 CTGGCCTGGTTTGTTTGTTTTTAA---ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGT 411
Db        414 ACGGCTGTAAAAAGTACCTTTTAAAGGAGAATAAAGCCAGGAGCAGTGGCTCACRTCTGT 473

Qy        412 GATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACC 471
Db        474 AATCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACC 533

Qy        472 AGCCAGGCCAACATGGTAAAACCATGTC 499
Db        534 AGCCTGGCCAACATGGCAAAACCCCATC 561


RESULT 3
US-10-027-632-139569
; Sequence 139569, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 139569
;   LENGTH: 624
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-139569

  Query Match           35.8%;  Score 179.2;  DB 13;  Length 624;
  Best Local Similarity   73.8%;  Pred. No. 2.9e-36;
  Matches  242;  Conservative    0;  Mismatches   83;  Indels    3;  Gaps    1;

Qy        175 TTGTTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTT 234
Db        235 TTTTTTGTGTGTTTTTTGTTTTGTTTTGCTTTTGTTTTTGAGACAGGGTCTCACTCTGCT 294

Qy        235 GCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTC 294
Db        295 GCCCAGGCTGAAGTGCAATGGGGTGATCATGGCTCACTGCAGTCTCAACCTTCTGGGCTC 354

Qy        295 AAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCAC 354
Db        355 AAGCAATCCTCCCACCTCAGCTTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCATGC 414

Qy        355 CTGGCCTGGTTTGTTTGTTTTTAA---ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGT 411
Db        415 ACGGCTGTAAAAAGTACCTTTTAAAGGAGAATAAAGCCAGGAGCAGTGGCTCACGTCTGT 474

Qy        412 GATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACC 471
Db        475 AATCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACCTGAGGTCAGGAGTTCGAGACC 534

Qy        472 AGCCAGGCCAACATGGTAAAACCATGTC 499
Db        535 AGCCTGGCCAACATGGCAAAACCCCATC 562


RESULT 4
US-10-017-161-1611
```

```
; Sequence 1611, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1611
;   LENGTH: 39725
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(39725)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(409)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1705)..(1966)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (3861)..(4281)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (6128)..(6513)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (13450)..(13658)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15063)..(15253)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (15939)..(16148)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (20042)..(20167)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22121)..(22419)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (22509)..(22680)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (23434)..(23601)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (24070)..(24306)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (27349)..(27643)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (28259)..(28404)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (28906)..(29142)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (30237)..(30376)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (36366)..(36636)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (36745)..(36988)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (39264)..(39525)
US-10-017-161-1611

  Query Match             35.1%;  Score 175.6;  DB 12;  Length 39725;
  Best Local Similarity   75.9%;  Pred. No. 1.5e-34;
  Matches  243;  Conservative    0;  Mismatches   74;  Indels    3;  Gaps    2;

Qy        182 TTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGG 241

Db       7409 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTGTCA-CCAGC 7467

Qy        242 CTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAAT 301

Db       7468 CTGGAGTGCAGTGGCATGATCTTGGCTCACTGCAACCTTTGCCTCCTATGTTCAAGTGAT 7527

Qy        302 CCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCT 361

Db       7528 TTTCCTGCCTCAGCCTCCTGAGTAACTGGGATTACAGGCATGAGCCACCGCGGCCGGCCA 7587

Qy        362 GGTTTGTTTGTTTTTA--AATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAG 419

Db       7588 GATTTGCATTTTTGAAACAACTGCTAGGCTGGGCGCGGTGGCTCACACCTGTAATCCCAG 7647

Qy        420 CACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGC 479

Db       7648 CACTGTGGGAGGCCGAGGCAGGTGGATCACCTGAGGTCAGGGGTTCGAGACCAGCCTGGC 7707

Qy        480 CAACATGGTAAAACCATGTC 499

Db       7708 CAACATGGTGAAACCCCGTC 7727


RESULT 5
US-10-017-161-1849/c
; Sequence 1849, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
;  APPLICANT: SUWA, MAKIKO
;  APPLICANT:  ASAI, KIYOSHI
;  APPLICANT:  AKIYAMA, YUTAKA
;  APPLICANT:  ABURATANI, HIROYUKI
;  TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
;  FILE REFERENCE: 084335/0152
;  CURRENT APPLICATION NUMBER: US/10/017,161
;  CURRENT FILING DATE:  2002-12-18
;  PRIOR APPLICATION NUMBER: JP 2001/246789
;  PRIOR FILING DATE: 2001-06-18
;  NUMBER OF SEQ ID NOS: 2430
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1849
;   LENGTH: 871
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: source
;   LOCATION: (1)..(871)
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (201)..(671)
US-10-017-161-1849

  Query Match             34.0%;  Score 169.8;  DB 12;  Length 871;
  Best Local Similarity   72.8%;  Pred. No. 8.7e-34;
  Matches  219;  Conservative    0;  Mismatches   82;  Indels    0;  Gaps    0;

Qy        199 TTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCAT 258

Db        821 TGTATTTTTTTTTAAGAGTCAGGGTCTTGCTCTGTTGCCCAGGCTAGAATGCAGTGGTGT 762
```

```
Qy        259 GATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTT 319
Db        761 GACCTTGGCTTACTACAGCCTAAAGCTCTTGGGCTCAAGTGATCCTCCTGCCTCAGCCTC 702

Qy        319 CCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAA 378
Db        701 CTGAGATGCTCGGAGTACAGGCATAAACTACTACAGCTGGCACACACTGTATTTAAGAGT 642

Qy        379 ATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGC 438
Db        641 AAAGGGGGGCCAGGCGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGC 582

Qy        439 AGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGT 498
Db        581 AGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGAGAAGCCCTGT 522

Qy        499 C 499
Db        521 C 521


RESULT 6
US-10-225-810-10/c
; Sequence 10, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
;  APPLICANT: Bermingham, Jr., John R.
;  TITLE OF INVENTION: Tramdorins and Methods of Using Tramdorin
;  FILE REFERENCE: McLaugh-07165
;  CURRENT APPLICATION NUMBER: US/10/225,810
;  CURRENT FILING DATE:  2002-08-21
;  NUMBER OF SEQ ID NOS: 76
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 10
;   LENGTH: 36400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-225-810-10

  Query Match           33.3%;  Score 166.6;  DB 12;  Length 36400;
  Best Local Similarity   71.7%;  Pred. No. 3e-32;
  Matches  236;  Conservative    0;  Mismatches   84;  Indels    9;  Gaps    1;

Qy        180 TTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCA 239
Db       3031 TTATGCATTTATTATTTTATTTTATTTTTATTTTTTGAGACAGGGTCTCACTGTCACCCA 2972

Qy        240 GGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCA 299
Db       2971 GGCTGAAGTGCAGTGGCATAATCACAGTTCACTGTGACCTCCTCCTCCCAGCCTCAAGCA 2912

Qy        300 ATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGC 359
Db       2911 ATCCTCCAACCTCAGCCTCCTGAGTAGTTGGGACTACAGGCATTCACCATCATGCCCAGC 2852

Qy        360 CTGGTTTGTTTGTTTTTAAA---------TTTTGAGGCCAGGTGCAGTGGCCCATATCTG 410
Db       2851 TAATTTTTGTATTTCTTGTAGATATGGGGTTTTGTGGCCGGGTGCAGTGGCTCATGCCTG 2792

Qy        411 TGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGAC 470
Db       2791 TAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTAGAGAC 2732

Qy        471 CAGCCAGGCCAACATGGTAAAACCATGTC 499
Db       2731 CAGCCTGGCCAACATGGTGAAACCCCATC 2703


RESULT 7
US-10-225-810-26/c
; Sequence 26, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
;  APPLICANT: Bermingham, Jr., John R.
;  TITLE OF INVENTION: Tramdorins and Methods of Using Tramdorin
;  FILE REFERENCE: McLaugh-07165
;  CURRENT APPLICATION NUMBER: US/10/225,810
;  CURRENT FILING DATE:  2002-08-21
;  NUMBER OF SEQ ID NOS: 76
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
;   LENGTH: 250000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (33774)..(33774)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (42953)..(43052)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45557)..(45656)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (48203)..(48302)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (49551)..(49650)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51561)..(51660)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (52722)..(52821)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (53864)..(53963)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (55290)..(55389)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56674)..(56773)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (57879)..(57978)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78952)..(79051)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85316)..(85415)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-225-810-26

  Query Match           33.3%;  Score 166.6;  DB 12;  Length 250000;
  Best Local Similarity   71.7%;  Pred. No. 7.1e-32;
  Matches  236;  Conservative    0;  Mismatches   84;  Indels    9;  Gaps    1;

Qy        180 TTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCA 239
Db     211200 TTATGCATTTATTATTTTATTTTATTTTTATTTTTTGAGACAGGGTCTCACTGTCACCCA 211141
```

```
Qy        240 GGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCA 299
Db     211140 GGCTGAAGTGCAGTGGCATAATCACAGTTCACTGTGACCTCCTCCTCCCAGCCTCAAGCA 211081

Qy        300 ATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGC 359
Db     211080 ATCCTCCAACCTCAGCCTCCTGAGTAGTTGGGACTACAGGCATTCACCATCATGCCCAGC 211021

Qy        360 CTGGTTTGTTTGTTTTTAAA---------TTTTGAGGCCAGGTGCAGTGGCCCATATCTG 410
Db     211020 TAATTTTTGTATTTCTTGTAGATATGGGGTTTTGTGGCCGGGTGCAGTGGCTCATGCCTG 210961

Qy        411 TGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGAC 470
Db     210960 TAATCCCAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTAGAGAC 210901

Qy        471 CAGCCAGGCCAACATGGTAAAACCATGTC 499
Db     210900 CAGCCTGGCCAACATGGTGAAACCCCATC 210872
```

RESULT 8
US-09-764-855-328
; Sequence 328, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 328
;   LENGTH: 32152
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-855-328

```
  Query Match             33.2%;  Score 165.8;  DB 10;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251
Db       3262 TGTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 3321

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
Db       3322 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 3381

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
Db       3382 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 3441

Qy        372 TTTTTAAA------------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
Db       3442 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 3501

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
Db       3502 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 3559

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
Db       3560 ACCTGGCCAACATGGTGAAACCCTGTC 3586
```

RESULT 9
US-09-764-872-518/c
; Sequence 518, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 518
;   LENGTH: 32152
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-518

```
  Query Match             33.2%;  Score 165.8;  DB 11;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251
Db      28891 TGTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 28832

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
Db      28831 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 28772

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
Db      28771 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 28712

Qy        372 TTTTTAAA------------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
Db      28711 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 28652

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
Db      28651 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 28594

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
Db      28593 ACCTGGCCAACATGGTGAAACCCTGTC 28567
```

RESULT 10
US-10-072-349-328
; Sequence 328, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; CURRENT FILING DATE: 2002-02-11
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 328
;   LENGTH: 32152
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-072-349-328

```
  Query Match             33.2%;  Score 165.8;  DB 14;  Length 32152;
  Best Local Similarity   74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251
Db       3262 TGTTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 3321

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311
```

```
Db       3322 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 3381

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371

Db       3382 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 3441

Qy        372 TTTTTAAA---------- -------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412

Db       3442 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 3501

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472

Db       3502 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 3559

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499

Db       3560 ACCTGGCCAACATGGTGAAACCCTGTC 3586


RESULT 11
US-09-764-855-327
; Sequence 327, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA110
;  CURRENT APPLICATION NUMBER: US/09/764,855
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
;   LENGTH: 32204
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-855-327

  Query Match           33.2%;  Score 165.8;  DB 10;  Length 32204;
  Best Local Similarity  74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251

Db       8128 TGTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 8187

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311

Db       8188 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 8247

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371

Db       8248 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 8307

Qy        372 TTTTTAAA----------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412

Db       8308 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 8367

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472

Db       8368 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 8425

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499

Db       8426 ACCTGGCCAACATGGTGAAACCCTGTC 8452


RESULT 12
US-09-764-872-517/c
; Sequence 517, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA125
;  CURRENT APPLICATION NUMBER: US/09/764,872
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 957
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 517
;   LENGTH: 32204
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-872-517

  Query Match           33.2%;  Score 165.8;  DB 11;  Length 32204;
  Best Local Similarity  74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251

Db      24077 TGTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 24016

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311

Db      24017 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 23958

Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371

Db      23957 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 23898

Qy        372 TTTTTAAA----------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412

Db      23897 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 23838

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472

Db      23837 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 23780

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499

Db      23779 ACCTGGCCAACATGGTGAAACCCTGTC 23753


RESULT 13
US-10-072-349-327
; Sequence 327, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA110C1
;  CURRENT APPLICATION NUMBER: US/10/072,349
;  CURRENT FILING DATE:  2002-02-11
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 327
;   LENGTH: 32204
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-072-349-327

  Query Match           33.2%;  Score 165.8;  DB 14;  Length 32204;
  Best Local Similarity  74.6%;  Pred. No. 4.6e-32;
  Matches  244;  Conservative    0;  Mismatches   62;  Indels   21;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCA 251

Db       8128 TGTTTTTTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTTGCCCAGGCTGGAGTGCC 8187

Qy        252 ATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCT 311

Db       8188 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCCTTCCAGGTTCATGCCATTCTCCTGCCT 8247
```

```
Qy        312 CAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTG 371
Db       8248 CAGCCTCCCGAGTAGCTGGGACTACAGGCGCCTGCCACCACACCCGGCTACAAAGGTGTA 8307

Qy        372 TTTTTAAA------------------TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTG 412
Db       8308 TTTTTTAAGTGACATAAAAATAAATGTCTTTGGGGCCAGGCGCAGTGGCTCACACCTGTA 8367

Qy        413 ATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCA 472
Db       8368 ATCCCAGCACTTTGGGAGGCCAAGGTGGGTGAAT--CATGAGGTCAGGAGTTCGAGACCA 8425

Qy        473 GCCAGGCCAACATGGTAAAACCATGTC 499
Db       8426 ACCTGGCCAACATGGTGAAACCCTGTC 8452


RESULT 14
US-10-027-632-276723
; Sequence 276723, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 276723
;   LENGTH: 662
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-276723

  Query Match             32.4%;  Score 162.2;  DB 13;  Length 662;
  Best Local Similarity   71.3%;  Pred. No. 6.9e-32;
  Matches  229;  Conservative    1;  Mismatches   84;  Indels    7;  Gaps    1;

Qy        174 TTTGTTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGT 233
Db        319 TTTTCTTTTCCTTTCTTTCTTTCTTTTTTAAATTTTTTAAGAGATAGGGTCTCACTCTGT 378

Qy        234 TGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTT 293
Db        379 CATCCAGGCTAGAGCACAGTGGCGTGATCACAGCTCACTGCAGCTTTAACCTCCCAGGCT 438

Qy        294 CAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCA 353
Db        439 CAAGCGATCCTCCTGCTTCAGCCTCCCAAGTAGCTGAGACTACAGGCATGTRCCACCACA 498

Qy        354 CCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGA 413
Db        499 CCCAGCTAATTTTG-------AAACTTTTATAGGCCAGGCATGGTGGCTCACGCCTGTCA 551

Qy        414 TCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAG 473
Db        552 TCCCAGCATTTTGGGAGGCTGAGGTGGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAG 611

Qy        474 CCAGGCCAACATGGTAAAACC 494
Db        612 CCTGGCCAACATGGAGAAACC 632


RESULT 15
US-10-027-632-147363/c
; Sequence 147363, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 147363
;   LENGTH: 798
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-147363

  Query Match             32.4%;  Score 162.2;  DB 13;  Length 798;
  Best Local Similarity   71.3%;  Pred. No. 7.5e-32;
  Matches  229;  Conservative    1;  Mismatches   84;  Indels    7;  Gaps    1;

Qy        174 TTTGTTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGT 233
Db        346 TTTTCTTTTCCTTTCTTTCTTTCTTTTTTAAATTTTTTAAGAGATAGGGTCTCACTCTGT 287

Qy        234 TGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTT 293
Db        286 CATCCAGGCTAGAGCACAGTGGCGTGATCACAGCTCACTGCAGCTTTAACCTCCCAGGCT 227

Qy        294 CAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCA 353
Db        226 CAAGCGATCCTCCTGCTTCAGCCTCCCAAGTAGCTGAGACTACAGGCATGTGCCACCACA 167

Qy        354 CCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGA 413
Db        166 CCCAGCTAATTTTG-------AAACTTTTATAGGCCAGGCATGGTGGCTCACGCCTGTCA 114

Qy        414 TCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAG 473
Db        113 TCCCAGCATTTTGGGAGGCTGAGGTGGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAG 54

Qy        474 CCAGGCCAACATGGTAAAACC 494
Db         53 CCTSGCCAACATGGAGAAACC 33


Search completed: October 24, 2003, 18:34:15
Job time : 127.169 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 1044.61 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: 1 taactgactctgctgtgttt.....aacatggtaaaaccatgtcg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	177.6	35.5	325	9	AI343123	AI343123 tb04e07.x
c	2	171.2	34.2	371	9	AW821209	AW821209 PM2-ST030
c	3	171.2	34.2	376	9	AW821105	AW821105 PM2-ST030
c	4	165.6	33.1	1027	10	BG622059	BG622059 602646665
	5	165.4	33.1	439	12	BI467370	BI467370 ic23e02.x
	6	164.2	32.8	824	28	AQ888705	AQ888705 HS_3144_A
c	7	163.4	32.7	359	13	BQ353356	BQ353356 IL3-HT061
	8	161.8	32.4	473	9	AA443390	AA443390 zw87a09.r
	9	157.8	31.6	681	29	AG059079	AG059079 Pan troglodytes
c	10	152	30.4	445	9	AA573381	AA573381 nm53a04.s
	11	151.2	30.2	391	13	BQ574362	BQ574362 UI-H-EZ1-
	12	150.8	30.2	680	12	BM990649	BM990649 UI-H-DI0-
	13	148.8	29.8	347	14	CB266516	CB266516 1005422 H
c	14	148.8	29.6	661	28	AQ539566	AQ539566 RPCI-11-3
	15	148.2	29.6	523	28	B69012	B69012 CIT-HSP-202
	16	147.8	29.6	585	29	AG016069	AG016069 Homo sapi
c	17	146.6	29.3	416	13	BU685269	BU685269 UI-CF-EU1
c	18	146.6	29.3	418	13	BU679484	BU679484 UI-CF-CU1
	19	146.6	29.3	1004	13	BX407955	BX407955 BX407955
	20	146	29.2	418	10	BF931715	BF931715 CM2-NT016
c	21	145.6	29.1	317	10	BE156651	BE156651 QV0-HT036
	22	145	29.0	367	12	BM682825	BM682825 UI-E-EJ1-
	23	145	29.0	436	9	AA649183	AA649183 ns31a11.s
	24	144.6	28.9	492	28	B45413	B45413 HS-1061-A1-
c	25	144.2	28.8	627	29	AG090681	AG090681 Pan troglodytes
c	26	143.8	28.8	479	9	AI863180	AI863180 tz44c04.x
	27	143.6	28.7	1052	13	BX463038	BX463038 BX463038
c	28	143	28.6	324	12	BM982363	BM982363 UI-CF-EN1
c	29	142.8	28.6	353	9	AI815608	AI815608 au43b01.y
	30	142.6	28.5	441	2	HSM085930	Bx493102 Homo sapi
	31	142.6	28.5	814	13	BU852129	BU852129 AGENCOURT
	32	142.4	28.5	502	2	HSM071632	Bx481465 Homo sapi
c	33	142.2	28.4	475	9	AI261932	AI261932 qk15g05.x
c	34	141.8	28.4	331	9	AL135357	AL135357 DKFZp762E
c	35	141.8	28.4	639	9	AW163068	AW163068 au91h05.y
c	36	141.8	28.4	729	9	AV703572	AV703572 AV703572
c	37	141.6	28.3	711	29	AG011827	AG011827 Homo sapi
c	38	141.2	28.2	508	12	BM991594	BM991594 UI-H-DF1-
	39	141	28.2	787	29	BZ611164	BZ611164 WHABT567F
	40	140.6	28.1	360	9	AA847961	AA847961 od72b11.s
	41	140.6	28.1	843	13	BU520971	BU520971 AGENCOURT
c	42	140.4	28.1	404	14	CD251403	CD251403 AGENCOURT
	43	140.2	28.0	390	9	AI418927	AI418927 tf52f03.x
c	44	139.8	28.0	827	12	BI457063	BI457063 603185984
c	45	139.8	28.0	1379	11	BC022205	BC022205 Homo sapi

ALIGNMENTS

RESULT 1
AI343123/c
LOCUS AI343123 325 bp mRNA linear EST 16-FEB-1999
DEFINITION tb04e07.x2 NCI_CGAP_Lu26 Homo sapiens cDNA clone IMAGE:2052612 3' similar to contains Alu repetitive element;contains element PTR5 repetitive element ;, mRNA sequence.
ACCESSION AI343123
VERSION AI343123.1 GI:4060329
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 325)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 437 Std Error: 0.00

Seq primer: -40UP from Gibco
High quality sequence stop: 293.

```
FEATURES             Location/Qualifiers
     source          1..325
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2052612"
                     /tissue_type="invasive adenocarcinoma"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Lu26"
                     /note="Organ: lung; Vector: pAMP1; mRNA made from lung
                     adenocarcinoma tissue, cDNA made by oligo-dT priming.
                     Directionally cloned. Size-selected on agarose gel,
                     average insert size 500 bp. Primary library,
                     non-amplified. "
BASE COUNT      100 a      81 c      89 g      54 t       1 others
ORIGIN
```

Query Match 35.5%; Score 177.6; DB 9; Length 325;
Best Local Similarity 75.1%; Pred. No. 8.3e+02;
Matches 235; Conservative 0; Mismatches 75; Indels 3; Gaps 1;

```
Qy        190 TTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATG 249
Db        325 TTTTTTTTTTTTTTTTTTTTTTTTGAGGCGGGGGCTCGCTCTGTTGCCCAGGCTGGAGTG 266

Qy        250 CAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGC 309
Db        265 CAGTGGCACAATCTTGGCTTACTGCAGCCTCCTCCTCCCGGGTTCAAGTGATTCTCCTGC 206

Qy        310 CTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTT 369
Db        205 CTCAGCCTCCTGGGTAGCTGGGACTATAGGTGTGAGCCACCACGCCCAGCCAGATTTAAT 146

Qy        370 T---GTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTG 426
Db        145 TTCATTTATTGCTCCTTTTTGCCGGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTG 86

Qy        427 GGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATG 486
Db         85 GGAGGCCAAAGTGGGCGGATCACCTGAGGTCAGGAGTTCCAGACCAGCCTGGCTAACATA 26

Qy        487 GTAAAACCATGTC 499
Db         25 GCAAAACCCCTTC 13
```

```
RESULT 2
AW821209/c
LOCUS       AW821209                371 bp    mRNA    linear   EST 17-MAY-2000
DEFINITION  PM2-ST0303-230200-006-g12 ST0303 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW821209
VERSION     AW821209.1  GI:7914203
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 371)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
```

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM2-ST0303-230
200-006-g12&t3=2000-02-23&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 12
High quality sequence stop: 371.

```
FEATURES             Location/Qualifiers
     source          1..371
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0303"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from CRESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      101 a      86 c      97 g      87 t
ORIGIN
```

Query Match 34.2%; Score 171.2; DB 9; Length 371;
Best Local Similarity 75.3%; Pred. No. 1.2e+03;
Matches 226; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

```
Qy        200 TTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATG 259
Db        358 TTTATTTATTTTTTGAGACAGGGCCTCACTCTGTTGTCCAGGCTGGAGTGCAGTGGTGCA 299

Qy        260 ATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTC 319
Db        298 GTCATAGGTCACTGCAACCTTGACCTCCTGGGCTCAAGAAATCCTCCCACCTCAGCCTCC 239

Qy        320 CAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAA 379
Db        238 CAAGTTGCTGGGACTACAGATGAATGCCACCATGTTTGG-CTGATTTTTACAATTTTAAA 180

Qy        380 TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCA 439
Db        179 TTTTTTGCCTGGGCGCAGTGGCTCATGCCTGTAATCTCAGCACTTTGGGAGGCCGAGGCG 120

Qy        440 GGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
Db        119 GGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCAAAACCCTGTC 60
```

```
RESULT 3
AW821105/c
LOCUS       AW821105                376 bp    mRNA    linear   EST 17-MAY-2000
DEFINITION  PM2-ST0303-281299-002-H05 ST0303 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW821105
VERSION     AW821105.1  GI:7914099
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 376)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
```

```
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=PM2-ST0303-281
            299-002-h05&t3=1999-12-28&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 376.
FEATURES             Location/Qualifiers
     source          1..376
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="ST0303"
                     /note="Organ: stomach; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  CRESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       99 a     92 c     97 g     88 t
ORIGIN

  Query Match             34.2%;  Score 171.2;  DB 9;  Length 376;
  Best Local Similarity   75.3%;  Pred. No. 1.2e+03;
  Matches  226;  Conservative    0;  Mismatches   73;  Indels    1;  Gaps    1;

Qy        200 TTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATG 259

Db        363 TTTATTTATTTTTTGAGACAGGGCCTCACTCTGTTGTCCAGGCTGGAGTGCAGTGGTGCA 304

Qy        260 ATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTC 319

Db        303 GTCATAGGTCACTGCAACCTTGACCTCCTGGGCTCAAGAAATCCTCCCACCTCAGCCTCC 244

Qy        320 CAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAA 379

Db        243 CAAGTTGCTGGGACTACAGATGAATGCCACCATGTTTGG-CTGATTTTTACAATTTTAAA 185

Qy        380 TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCA 439

Db        184 TTTTTTGCCTGGGCGCAGTGGCTCATGCCTGTAATCTCAGCACTTTGGGAGGCCGAGGCG 125

Qy        440 GGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499

Db        124 GGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCAAAACCCTGTC 65


RESULT 4
BG622059/c
LOCUS       BG622059                1027 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646665F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4768049 5',
            mRNA sequence.
ACCESSION   BG622059
VERSION     BG622059.1  GI:13673430
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1027)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1631  row: b  column: 18
            High quality sequence stop: 444.
FEATURES             Location/Qualifiers
     source          1..1027
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4768049"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_79"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      336 a    231 c    318 g    142 t
ORIGIN

  Query Match             33.1%;  Score 165.6;  DB 10;  Length 1027;
  Best Local Similarity   73.4%;  Pred. No. 8.5e-02;
  Matches  226;  Conservative    0;  Mismatches   79;  Indels    3;  Gaps    1;

Qy        186 TTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGG 245

Db        444 TTTTTTTTTTTTTTTTTTTTTTTTTTTGCGATGGAGTCTCACTCTATTGCCCAGGCTGG 385

Qy        246 AATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTC 305

Db        384 AATACAGTGGCACAATCTTGGTTCACTGCAATCTCCACATCCCGGGCTCAAGCGATCCTC 325

Qy        306 CTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTT 365

Db        324 CTGCCTCAGCCTCCCGAATAGCTGGGATTACAAGAGTGCATCACCAGGCCT---AATTTT 268

Qy        366 TGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTT 425

Db        267 TGTATTTTTAGTTGAGACAGGGCCAGGCGTGGTGGCTCACGCCTATAATCTCAGCACTTT 208

Qy        426 GGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACAT 485

Db        207 GGGAGGCTGAGGCAGGTGGATCACTTGAGGCCAGGAGTTCGAGACCAGCCTGGTCAACAT 148

Qy        486 GGTAAAAC 493

Db        147 GATGAAAC 140


RESULT 5
BI467370
LOCUS       BI467370                 439 bp    mRNA    linear   EST 22-AUG-2001
DEFINITION  ic23e02.x3 HR85 islet Homo sapiens cDNA 3', mRNA sequence.
ACCESSION   BI467370
```

```
VERSION     BI467370.1  GI:15280248
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 439)
  AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
            ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
            , Jackson,Y. and Bowers,Y.
  TITLE     Endocrine Pancreas Consortium
  JOURNAL   Unpublished
COMMENT     Other_ESTs: ic23e02.y3
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge.
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            High quality sequence stop: 439.
FEATURES             Location/Qualifiers
     source          1..439
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="Purified pancreatic islet"
                     /lab_host="DH10B"
                     /clone_lib="HR85 islet"
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
BASE COUNT      102 a    113 c    107 g    116 t      1 others
ORIGIN

  Query Match             33.1%;  Score 165.4;  DB 12;  Length 439;
  Best Local Similarity   74.0%;  Pred. No. 1.6e+03;
  Matches  222;  Conservative    0;  Mismatches   77;  Indels    1;  Gaps    1;

Qy        200 TTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAATGGCATG 259
Db         76 TTTATTTTATTTTTGAGACAGGGCCTCACTCCGTCGTCCAGGCTGGAGTGCAGTGGCGCA 135

Qy        260 ATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCAATCTTC 319
Db        136 GTCACAGGTCACCGCAACCTTGACCTCCTGGGCTCAAGAAATCCTCCCACCTCAGCCTCC 195

Qy        320 CAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTTTTTAAA 379
Db        196 CAAGTTGCTGGGACCACAGATGAATGCCACCATGTTGG-CTGATTTTTACAATTTTAAA 254

Qy        380 TTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACCAAGGCA 439
Db        255 TTTTTGCCTGGGCGCAGTGGCTCATGCCTGTAATCTCANCACTTTGGGAGGCTGAGGCG 314

Qy        440 GGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAACCATGTC 499
Db        315 GGCAGATCACCTGAGGTCAGGAGTTCAAGACCAGCCTGGCCAACATGGCAAAACCTTGTC 374
```

```
RESULT 6
AQ888705
LOCUS       AQ888705                 824 bp    DNA     linear   GSS 10-NOV-1999
DEFINITION  HS_3144_A2_E05_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3144 Col=10 Row=I, genomic survey
            sequence.
ACCESSION   AQ888705
VERSION     AQ888705.1  GI:6344895
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 824)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3616
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3144  row: I  column: 10
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 824.
FEATURES             Location/Qualifiers
     source          1..824
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=3144 Col=10 Row=I"
                     /sex="male"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      199 a    113 c    298 g    213 t      1 others
ORIGIN

  Query Match             32.8%;  Score 164.2;  DB 28;  Length 824;
  Best Local Similarity   70.8%;  Pred. No. 1.1e+03;
  Matches  235;  Conservative    0;  Mismatches   89;  Indels    8;  Gaps    1;

Qy        176 TGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTG 235
Db         18 TGGTGTTTGAAGCAACTACTTAAATTTTTTTATTTTTGAGACGGAGTCTTGCTCTGTCA 77

Qy        236 CCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCA 295
Db         78 TCCAGGCTGGCATGCAATGGCGCAATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTTA 137

Qy        296 AGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACC 355
Db        138 AGCAATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGTGCACGCCACTATGCC 197

Qy        356 TGGCCTGGTTTGTTTGT--------TTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATAT 407
Db        198 CGGCTAATTTTTTGTATAATATATATGTATANATTTGGGTCGGGCACGGTGGCTCACAC 257

Qy        408 CTGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAA 467
Db        258 CTATAATCCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACCTGAGGTCAGGATGTAAA 317
```

```
Qy        468 GACCAGCCAGGCCAACATGGTAAAACCATGTC 499

Db        318 TACCAGCCTGACCAACATGGTGAAACCCCATC 349


RESULT 7
BQ353356/c
LOCUS       BQ353356                 359 bp    mRNA    linear   EST 20-MAY-2002
DEFINITION  IL3-HT0619-290800-252-B07 HT0619 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BQ353356
VERSION     BQ353356.1  GI:21017412
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-HT0619-
            290800-252-B07&t3=2000-08-29&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 4
            High quality sequence stop: 359.
FEATURES             Location/Qualifiers
     source          1..359
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0619"
                     /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      114 a     79 c     88 g     77 t      1 others
ORIGIN

  Query Match             32.7%;  Score 163.4;  DB 13;  Length 359;
  Best Local Similarity   75.1%;  Pred. No. 2.1e+03;
  Matches  217;  Conservative    0;  Mismatches   71;  Indels    1;  Gaps    1;

Qy        189 TTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAAT 248

Db        288 TCTTTTTTTTTTTTTTTTTTTTTTTGAGACAGGGTTTCACTCTGTCACCCAGGCTGGAGT 229

Qy        249 GCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTG 308

Db        228 GCAGTGGGGTGATCTTGGCTCACTGCTTCCTCAACCTCCCAGGTTCAAGCAGTCCTCCCA 169

Qy        309 CCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGT 368

Db        168 CCTCAACCTCCCAAGTAGCTGGGACTACAGGTGTGCGACACCACACC-GGGCTAAATATT 110

Qy        369 TTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGG 428

Db        109 CAAAAGTTTTTTGTAGGGCCGGGTGCAGTGGCTCACACCTGTAATCCCAGCACTTTGGG 50

Qy        429 AGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAG 477

Db         49 AGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTTGAGACCAGCCTG 1


RESULT 8
AA443390
LOCUS       AA443390                 473 bp    mRNA    linear   EST 03-JUN-1997
DEFINITION  zw87a09.r1 Soares_total_fetus_Nb2HF8_9w Homo sapiens cDNA clone
            IMAGE:783928 5' similar to contains Alu repetitive element;, mRNA
            sequence.
ACCESSION   AA443390
VERSION     AA443390.1  GI:2156065
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 473)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B.
            , Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie
            ,T., Waterston,R. and Wilson,R.
  TITLE     WashU-Merck EST Project 1997
  JOURNAL   Unpublished
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 430.
FEATURES             Location/Qualifiers
     source          1..473
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="GDB:5981546"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:783928"
                     /dev_stage="8-9 weeks"
                     /lab_host="DH10B"
                     /clone_lib="Soares_total_fetus_Nb2HF8_9w"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from pooled 8-9 week
                     (total) fetus material with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCTTAATTTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization, and was
                     constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      116 a    124 c    126 g    107 t
ORIGIN

  Query Match             32.4%;  Score 161.8;  DB 9;  Length 473;
  Best Local Similarity   73.8%;  Pred. No. 1.9e+03;
  Matches  234;  Conservative    0;  Mismatches   77;  Indels    6;  Gaps    2;

Qy        183 TTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGC 242

Db         46 TTTTTTTGTTTGTTTGTTTGTTTGTTTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGC 105
```

```
Qy        243 TGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATC 302
Db        106 TGGAGTGCAGTGGCGTGATCTCGGCTTACTGCAACCTCCGCCTCCCGGGTTCAAGCGATT 165

Qy        303 CTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTG 362
Db        166 CTCCTGCCTCAGCCTCCCAAGTAACTGGGACAACAGGCGCGTCCCACCACAC----CCAT 221

Qy        363 GTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCAC 422
Db        222 CTCTACTAAAAAATACAAAATTAGGCCGGGCGCGGTGGCTCATGCCTGTAATCCCAGCAC 281

Qy        423 TTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAA 482
Db        282 TTTGGGAGGCCGAGGCGGGCGGATCAC--GAGGTCAGAGATCGAGACCATCCTGGCTAA 339

Qy        483 CATGGTAAAACCATGTC 499
Db        340 CATGGTGAAACCCCATC 356
```

```
RESULT 9
AG059079
LOCUS       AG059079               681 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-046C15.R, genomic survey sequence.
ACCESSION   AG059079
VERSION     AG059079.1  GI:16610309
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 681)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                 Sequencing: M13Rev
             LIBRARY
                 Vector     : pKS145
                 R.Site 1   : SacI
                 R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..681
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-046C15.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      152 a    193 c    138 g    194 t      4 others
ORIGIN

  Query Match             31.6%;  Score 157.8;  DB 29;  Length 681;
  Best Local Similarity   73.3%;  Pred. No. 2e-03;
  Matches  231;  Conservative    0;  Mismatches   77;  Indels    7;  Gaps    2;

Qy        192 TATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTC-TGTTGCCCAGGCTGGAATGC 250
Db        212 TATAATTTTTTTTTTTTTTTTTAGACGGAGTTTCGCTCTTGTTGCCGGGGCTGGAGTGC 271

Qy        251 AATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCC 310
Db        272 AATGGTGCAATCTTGGCTCACTGCAACCTCCACCTTCCAGGTTCAAACGATTCTCCTGCC 331

Qy        311 TCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTT- 369
Db        332 TCAGCCTCCCAAATAGCTGGGATCACAGGCATGTGCCACCATGCCCAGCTAATTTTGTTA 391

Qy        370 -----TGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTT 424
Db        392 AAGTATAATTTTTAAAAATTATTACAGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTT 451

Qy        425 TGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACA 484
Db        452 TGAGAGGCCGAGGTGGGCAGATCGCTTAACGTCAAGAGTTTGAGACCAGCCTGGCCAACA 511

Qy        485 TGGTAAAACCATGTC 499
Db        512 TGGCGAAACCTCATC 526
```

```
RESULT 10
AA573381/c
LOCUS       AA573381               445 bp    mRNA    linear   EST 12-SEP-1997
DEFINITION  nm53a04.s1 NCI_CGAP_Br2 Homo sapiens cDNA clone IMAGE:1071918 3'
            similar to contains Alu repetitive element;contains L1.b2 L1
            repetitive element ;, mRNA sequence.
ACCESSION   AA573381
VERSION     AA573381.1  GI:2347909
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 445)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1466   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 411.
FEATURES             Location/Qualifiers
     source          1..445
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1071918"
                     /sex="female, pooled"
                     /tissue_type="breast"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Br2"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from pooled bulk
                     breast tumor tissue, and was then primed with a Not I -
                     oligo(dT) primer. Double-stranded cDNA was ligated to Eco
                     RI adaptors (Pharmacia), digested with Not I and cloned
                     into the Not I and Eco RI sites of the modified pT7T3
                     vector.  This library is the normalized version of
                     NCI_CGAP_Br1.1. Library was constructed by Bento Soares
                     and M. Fatima Bonaldo. "
```

```
BASE COUNT      137 a    112 c    102 g     94 t
ORIGIN

  Query Match             30.4%;  Score 152;  DB 9;  Length 445;
  Best Local Similarity   67.2%;  Pred. No. 4e+03;
  Matches  215;  Conservative    0;  Mismatches  105;  Indels    0;  Gaps    0;

Qy        169 TTGGTTTTGTTTTTTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGC 228

Db        385 TTTGTTATTCCTGCTTTCTGTCCATATATTTTTCTTTTTTTTAAGTGACAGGGTCTTGC 326

Qy        229 TCTGTTGCCCAGGCTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTC 288

Db        325 TCTGTTGTCCTGGCTGGAGTGCAATGGCACCATCATGGCTCACTGCAGCCTTGACCTCCC 266

Qy        289 GGGTTCAAGCAATCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTG 348

Db        265 AGGTTCAAGTAATCCTCCCATCTTGGCCTCCCGAGTAGCTGGGACTGCAGGCACACATCA 206

Qy        349 CCGCACCTGGCCTGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATC 408

Db        205 CCATACCCAGCTAAATTTAAAAAAAAAATTTTTTTTGGCCGGGTGCGGTGGCTCATGCC 146

Qy        409 TGTGATCCCAGCACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAG 468

Db        145 TGTAGTCCAAGCACTTTGGGAGGCCAAGGTGGGCAGATCACCTGAGGTCAGGAGTTGGAG 86

Qy        469 ACCAGCCAGGCCAACATGGT 488

Db         85 ACCTGGGAGGCGGAGGTTGT 66


RESULT 11
BQ574362
LOCUS       BQ574362                 391 bp    mRNA    linear   EST 19-JUN-2002
DEFINITION  UI-H-EZ1-baz-f-10-0-UI.s1 NCI_CGAP_Ch2 Homo sapiens cDNA clone
            UI-H-EZ1-baz-f-10-0-UI 3', mRNA sequence.
ACCESSION   BQ574362
VERSION     BQ574362.1  GI:21477679
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Steven Gitelis/ Rush Presbiterian, Dept. of
            Orthoapedics
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
             The following repetitive elements were found in this cDNA
            sequence: 11-171, >ALU (matched compliment) 192-390, >ALU
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..391
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-EZ1-baz-f-10-0-UI"
                     /tissue_type="Chondrosarcoma Grade II"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_Ch2"
                     /note="Organ: Left Pelvis; Vector: pT7T3-Pac (Pharmacia)
                     with a modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_Ch2 is a normalized cDNA library containing the
                     following tissue(s): Chondrosarcoma Grade II. The library
                     was constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     TGATCACGCT.
                     TAG_LIB=UI-H-EZ1
                     TAG_TISSUE=grade-2-chondrosarcoma
                     TAG_SEQ=ATCTAATATG"
BASE COUNT       93 a    100 c     91 g    107 t
ORIGIN

  Query Match             30.2%;  Score 151.2;  DB 13;  Length 391;
  Best Local Similarity   70.6%;  Pred. No. 4.6e+03;
  Matches  216;  Conservative    0;  Mismatches   88;  Indels    2;  Gaps    1;

Qy        194 TTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253

Db          1 TTTTTTTTTTTTTTTTTTAAGAGACAGGGTCTCACTGTGTCACCCAGACTGGAGTGCAGT 60

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313

Db         61 GGCATGATCACAGCTCACTGTTGCCTTGACCTCCTGGGCTCAAGTGATCCTTCCACCTCA 120

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTT 373

Db        121 GCCTCCCAAGTATTTGGGACCACATGCATGCATCACCACACCTGACAAATTAAAAAAAAA 180

Qy        374 TTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACC 433

Db        181 TTT--TTTTTTTTGCCAGGTGCAGTGGCTCACATCTGTAATCCCAGCACTCTGAGAGGCT 238

Qy        434 AAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAAC 493

Db        239 GAGGTGGGCGGATCACCTGAGGTTGGGAGTTTGAGACCAGCCTGACCAACACAGAGAAAC 298

Qy        494 CATGTC 499

Db        299 CCTGTC 304


RESULT 12
BM990649
LOCUS       BM990649                 689 bp    mRNA    linear   EST 17-JUN-2002
DEFINITION  UI-H-DI0-atp-c-13-0-UI.s1 NCI_CGAP_DI0 Homo sapiens cDNA clone
            IMAGE:5862756 3', mRNA sequence.
ACCESSION   BM990649
VERSION     BM990649.1  GI:19710038
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 689)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
```

```
            The following repetitive elements were found in this cDNA
            sequence: 11-168, >ALU (matched compliment) 199-418, >ALU 334-623,
            >ALU 634-661, >AT_rich#Low_complexity
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..689
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5862756"
                     /tissue_type="Lung Focal Fibrosis"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI_CGAP_DI0"
                     /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     NCI_CGAP_DI0 is a cDNA library containing the following
                     tissue(s): A pool of Lung Focal Fibrosis. The library was
                     constructed according to Bonaldo, Lennon and Soares,
                     Genome Research, 6:791-806, 1996. First strand cDNA
                     synthesis was primed with an oligo-dT primer containing a
                     Not I site. Double stranded cDNA was ligated to an EcoR I
                     adaptor, digested with Not I, and cloned directionally
                     into pT7T3-Pac vector. The oligonucleotide used to prime
                     the synthesis of first-strand cDNA contains a library tag
                     sequence that is located between the Not I site and the
                     (dT)18 tail. The sequence tag for this library is
                     ATACGCGGTC.
                     TAG_LIB=UI-H-DI0
                     TAG_TISSUE=lung with fibrosis
                     TAG_SEQ=ATACGCGGTC"
BASE COUNT      189 a    161 c    170 g    168 t      1 others
ORIGIN

  Query Match             30.2%;  Score 150.8;  DB 12;  Length 689;
  Best Local Similarity   70.2%;  Pred. No. 3.2e+02;
  Matches  219;  Conservative    0;  Mismatches   87;  Indels    6;  Gaps    1;

Qy        194 TTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253
Db          1 TTTTTTTTTTTTTTTTTTTAAGAGACAAGGTTTTGCTCTGCTGCCCAAGCTGGAGTGCAGT 60

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313
Db         61 GGCATGATCAGAGTTCACTGCAGCCTCAAACTCCTATGCTCAAGCAATTCTCTCACCTCA 120

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTG------ 367
Db        121 GTCTCCTGAGTAGCTAGGACTTCAGGCACGTGCCACCATGCCTGGCCACTAAAGAAAAAT 180

Qy        368 TTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGG 427
Db        181 ATTTTAAAAAATAACAGTGGCTGGGTACAGTGGCTCACGCCTGTAATCCCAGCACTTTGG 240

Qy        428 GAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGG 487
Db        241 GAGGCCGACGCAGGTGGATCACTTGAGGACAGCAGTTTGAAGCCAGCCTGGCCAACATGG 300

Qy        488 TAAAACCATGTC 499
Db        301 CATAACCTTGTC 312


RESULT 13
CB266516
LOCUS       CB266516                 347 bp    mRNA    linear   EST 20-FEB-2003
DEFINITION  1005422 Human Fat Cell 5'-Stretch Plus cDNA Library Homo sapiens
            cDNA 5', mRNA sequence.
ACCESSION   CB266516
VERSION     CB266516.1  GI:28441102
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 347)
  AUTHORS   Yang,R.-Z., Shuldiner,A. and Gong,D.-W.
  TITLE     EST analysis of human adipose gene expression
  JOURNAL   Unpublished
COMMENT     Contact: Gong Da-Wei
            Division of Endocrinology, Diabetes and Nutrition
            University of Maryland
            660 Redwood St, HH497, Baltimore, MD 21201, USA
            Tel: 410 706 1672
            Fax: 410 706 1622
            Email: dgong@medicine.umaryland.edu
            PCR PRimers
            FORWARD: CTCGGGAAGCGCGCCATTGTGTTGGT
            BACKWARD: AATACGACTCACTATAGGGCGAATTGG
            Seq primer: GTTGGTACCCGGGAATTC.
FEATURES             Location/Qualifiers
     source          1..347
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /sex="Male and Female"
                     /tissue_type="Adipose"
                     /clone_lib="Human Fat Cell 5'-Stretch Plus cDNA Library"
                     /note="Vector: lambdaTriplEx"
BASE COUNT       88 a     84 c     80 g     95 t
ORIGIN

  Query Match             29.8%;  Score 148.8;  DB 14;  Length 347;
  Best Local Similarity   70.7%;  Pred. No. 5.9e-03;
  Matches  212;  Conservative    0;  Mismatches   87;  Indels    1;  Gaps    1;

Qy        194 TTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCTGGAATGCAAT 253
Db         24 TATTATTATTATTATTTTTGAGACGGAGTCTCGTTCTGTCACCCAGGCTGGAGTGCAGT 83

Qy        254 GGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCCTCCTGCCTCA 313
Db         84 GGCACGATCTCAGCTCACTGCAAGCTTTGCCTCCTGGTTTCACGCCATTCTCCTGCCTCA 143

Qy        314 ATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCTGGTTTGTTTGTT 373
Db        144 GCCTCCCGAGTAGCTGGGACTAAAGACACCTGCTACCG-TCTTGCTATTATTAATAGAAA 202

Qy        374 TTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCACTTTGGGAGACC 433
Db        203 CATAACTGTAAGGCCAGGGCATGGTGGCTTATGCCTGTAGTCCCAGCACTTTGGGAGGCT 262

Qy        434 AAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCAACATGGTAAAAC 493
Db        263 GAGGTGGGAAGATTACTTGAGCCCAGGAATTCAAGACCAGCCTGGGCAACATAGCAAGAC 322


RESULT 14
AQ539566/c
LOCUS       AQ539566                 661 bp    DNA     linear   GSS 19-MAY-1999
DEFINITION  RPCI-11-338N10.TV RPCI-11 Homo sapiens genomic clone RPCI-11-338N10
            , genomic survey sequence.
ACCESSION   AQ539566
VERSION     AQ539566.1  GI:4870140
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
```

```
COMMENT     Other_GSSs: RPCI-11-338N10.TJ
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7629729"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-338N10"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT     141 a    202 c    148 g    170 t
ORIGIN

  Query Match             29.8%;  Score 148.8;  DB 28;  Length 661;
  Best Local Similarity   68.9%;  Pred. No. 3.8e+03;
  Matches  219;  Conservative    0;  Mismatches   97;  Indels    2;  Gaps    1;

Qy        184 TGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGGCT 243

Db        358 TGTGTTTTATTTCTATTATTATTATTTTTCGAGACAGGGCCTCACTCTGTTGCCCAGGCT 299

Qy        244 GGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCTTTCGGGTTCAAGCAATCC 303

Db        298 GCAGCCCAGTGGTGCCATCGCTGATCAGGCCAGCCTCGAACGCCTGGGCTCAAGCGATCC 239

Qy        304 TCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCCT-- 361

Db        238 TCCTGCCTCGGCCTCCCAAGGTGCTGGGATTACAGGCGTGAGCCACCATGGTCGGACTAA 179

Qy        362 GGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGCA 421

Db        178 ATTATGTTTAAAAGGCTCAGTCCGGGCTGGGCCTGGTGGCTTACACCTGTAATCCCAGAA 119

Qy        422 CTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCCA 481

Db        118 CTTTGGGAGGCCAAGGCAGGAGGATCACTTGAGATCAGGAGTTCGAGACCAGCCTGGCCA 59

Qy        482 ACATGGTAAAACCATGTC 499

Db         58 ACATGGTGAAACCCTGTC 41


RESULT 15
B69012
LOCUS       B69012                   523 bp    DNA     linear   GSS 21-JUN-1998
DEFINITION  CIT-HSP-2025G21.TF CIT-HSP Homo sapiens genomic clone 2025G21,
            genomic survey sequence.
ACCESSION   B69012
VERSION     B69012.1  GI:2667722
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 523)
  AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CIT-HSP-2025G21.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..523
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7046951"
                     /db_xref="taxon:9606"
                     /clone="2025G21"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT     127 a    127 c    134 g    135 t
ORIGIN

  Query Match             29.6%;  Score 148.2;  DB 28;  Length 523;
  Best Local Similarity   72.7%;  Pred. No. 4.6e+03;
  Matches  232;  Conservative    0;  Mismatches   83;  Indels    4;  Gaps    3;

Qy        182 TTTGTTGTTTTATTTTATTTTATTTTATTTTCTAGATGGAGTCTTGCTCTGTTGCCCAGG 241

Db         31 TTTTGTCTGTGAATTTTTGTTTGTTTGTTTTGAGACAAGGTCTTGCTCTGTTGCCCAAG 90

Qy        242 CTGGAATGCAATGGCATGATCTCAGCTCACTGCAACCTCCACCT-TTCGGGTTCAAGCAA 300

Db         91 CGGGAGTGCAGTAGCATGATCATAGCTCACTGCATCCTCGATATCCTAGGCTTCAGGTGA 150

Qy        301 TCCTCCTGCCTCAATCTTCCAAGTAGCTGGGACTACAGGAATGAGCTGCCGCACCTGGCC 360

Db        151 TCCTCCTGCCTCAGCCTCTCAAGTAGCTAGGACCACAAGCATGCGCCATCACACCT-AAC 209

Qy        361 TGGTTTGTTTGTTTTTAAATTTTGAGGCCAGGTGCAGTGGCCCATATCTGTGATCCCAGC 420

Db        210 TATTTTTTTTTTAATTTTAGTAGAGGCCAGGCGCAGTGGCTTACGCCTGTAATCCCAGC 269

Qy        421 ACTTTGGGAGACCAAGGCAGGCCGATTACTTGAGGTCAGGAGTTCAAGACCAGCCAGGCC 480

Db        270 ACTTTGGGAGGCCAAGGCGGGTGGAT--CATGAGGTCAGGAGTTCGAGACCAGCCTGGTC 327

Qy        481 AACATGGTAAAACCATGTC 499

Db        328 AACATGGTGAAACCCCATC 346


Search completed: October 24, 2003, 22:50:57
Job time : 1049.81 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1384.09 Seconds
(without alignments)
14809.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaaagtgctgggattacag.....tgaatgaggctagctaggcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		501	100.0	101584	9	CNS01DS5	AL121655 BAC seque
2		501	100.0	110000	6	AX093471	AX093471 Sequence
3		501	100.0	110000	9	HSA246003	AC246003 Homo sapi
4		501	100.0	155943	9	AC012364	AC012364 Homo sapi
5		98.8	19.7	127192	9	HS1065O2	AL035562 Human DNA
6		97.4	19.4	73036	2	AC091063	AC091063 Homo sapi
c 7		94	18.8	194835	9	AL158158	AL158158 Human DNA
c 8		93.6	18.7	142728	9	HSDJ792G4	AL049636 Human DNA
9		85	17.0	75046	9	BX248112	BX248112 Human DNA
10		85	17.0	150770	9	AL953911	AL953911 Human DNA
c 11		85	17.0	187035	2	BX088588	BX088588 Homo sapi
c 12		84.4	16.8	63409	2	AC022982	AC022982 Homo sapi
13		84.4	16.8	97811	9	AC008634	AC008634 Homo sapi
c 14		84.4	16.8	104475	9	AC011384	AC011384 Homo sapi
15		84.2	16.8	121394	2	AL137799	AL137799 Homo sapi
c 16		84.2	16.8	190570	2	AL157894	AL157894 Homo sapi
17		83.8	16.7	93418	9	AC008379	AC008379 Homo sapi
c 18		83.8	16.7	179087	2	AC069489	AC069489 Homo sapi
c 19		83.8	16.7	199952	2	AP002016	AP002016 Homo sapi
c 20		83.4	16.6	61840	2	AC130291	AC130291 Homo sapi
21		83.4	16.6	132383	9	BX296545	BX296545 Human DNA
22		83.4	16.6	197676	2	AC022544	AC022544 Homo sapi
c 23		82.6	16.5	126613	9	HSJ603I14	AL122001 Human DNA
24		82.2	16.4	163709	2	AP001155	AP001155 Homo sapi
c 25		82.2	16.4	185278	9	AC023043	AC023043 Homo sapi
c 26		82.2	16.4	195489	2	AP001159	AP001159 Homo sapi
27		82.2	16.4	210624	2	AP001321	AP001321 Homo sapi
c 28		81.8	16.3	84506	9	AL591867	AL591867 Human DNA
c 29		79.6	15.9	38476	9	U73024	U73024 Homo sapien
c 30		79.6	15.9	242825	9	HSPEX	Y10196 Homo sapien
31		78.2	15.6	137718	9	AL137066	AL137066 Human DNA
32		77.6	15.5	68562	2	AC104936	AC104936 Homo sapi
33		77.6	15.5	154893	2	AC072038	AC072038 Homo sapi
c 34		77.6	15.5	167108	9	CNS01RHA	AL161666 Human chr
c 35		77.6	15.5	193486	9	AC012123	AC012123 Homo sapi
c 36		77.4	15.4	89229	2	AC016273	AC016273 Homo sapi
c 37		77	15.4	162199	9	AC010096	AC010096 Homo sapi
c 38		77	15.4	178671	2	AC069283	AC069283 Homo sapi
c 39		76	15.2	101006	2	AC079392	AC079392 Homo sapi
40		76	15.2	111073	9	AC006375	AC006375 Homo sapi
c 41		75.8	15.1	100122	2	AF178220	AF178220 Homo sapi
42		75.8	15.1	123331	9	AC105416	AC105416 Homo sapi
c 43		75.8	15.1	155797	2	AC073049	AC073049 Homo sapi
c 44		75.8	15.1	157606	9	AC090897	AC090897 Homo sapi
45		75.8	15.1	194498	2	AC012359	AC012359 Homo sapi

ALIGNMENTS

RESULT 1
CNS01DS5
LOCUS CNS01DS5 101584 bp DNA linear PRI 18-APR-2002
DEFINITION BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 336P14 of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION AL121655
VERSION AL121655.1 GI:6002388
KEYWORDS SPG4 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 101584)
AUTHORS Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D., Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,

```
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a novel AAA protein, is altered in the most frequent form
            of autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. (1999) In press
REFERENCE   2  (bases 1 to 101584)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1..101584
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="336P14"
                     /clone_lib="CITB_978_SKB"
BASE COUNT    29860 a  22110 c  22047 g  27567 t
ORIGIN

  Query Match            100.0%;  Score 501;  DB 9;  Length 101584;
  Best Local Similarity  100.0%;  Pred. No. 4.9e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db      63973 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 64032

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
Db      64033 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 64092

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 180
Db      64093 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 64152

Qy        181 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 240
Db      64153 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 64212

Qy        241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 300
Db      64213 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 64272

Qy        301 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db      64273 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 64332

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 420
Db      64333 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 64392

Qy        421 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 480
Db      64393 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 64452

Qy        481 ATGAATGAGGCTAGCTAGGCC 501
Db      64453 ATGAATGAGGCTAGCTAGGCC 64473

RESULT 2
AX093471
LOCUS       AX093471               110000 bp    DNA     linear   PAT 30-MAR-2001
DEFINITION  Sequence 1 from Patent WO0118198.
ACCESSION   AX093471
VERSION     AX093471.1  GI:13509911
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Weissenbach,J. and Hazan,J.
  TITLE     Cloning, expression and characterisation of the spg4 gene
            responsible for the most frequent form of autosomal spastic
            paraplegia
  JOURNAL   Patent: WO 0118198-A 1 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1..110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     exon            9932..10471
     intron          10472..33718
     exon            33719..33805
     intron          33806..35748
     exon            35749..35832
     intron          35833..45022
     exon            45023..45118
     intron          45119..60863
     exon            60864..61051
     intron          61052..61927
     exon            61928..62061
     intron          62062..62344
     exon            62345..62438
     intron          62439..73173
     exon            73174..73248
     intron          73249..74633
     exon            74634..74705
     intron          74706..82788
     exon            82789..82864
     intron          82865..83102
     exon            83103..83194
     intron          83195..83334
     exon            83335..83414
     intron          83415..88129
     exon            88130..88172
     intron          88173..89561
     exon            89562..89641
     intron          89642..91162
     exon            91163..91233
     intron          91234..93443
     exon            93444..93484
     intron          93485..100599
     exon            100600..102009
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match            100.0%;  Score 501;  DB 6;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db       4000 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 4059

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
Db       4060 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 4119

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 180
Db       4120 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 4179

Qy        181 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 240
Db       4180 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 4239

Qy        241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 300
Db       4240 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 4299
```

```
Qy        301 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       4300 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 4359

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 420
Db       4360 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 4419

Qy        421 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 480
Db       4420 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 4479

Qy        481 ATGAATGAGGCTAGCTAGGCC 501
Db       4480 ATGAATGAGGCTAGCTAGGCC 4500


RESULT 3
HSA246003
LOCUS       HSA246003             110000 bp    DNA     linear   PRI 08-MAR-2000
DEFINITION  Homo sapiens Spast gene for spastin protein.
ACCESSION   AJ246003
VERSION     AJ246003.1  GI:6273492
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 110000)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The
            sequence is the result of the assembly of 2 BAC clones: R-336p14
            and
            563N4, respectively from RPCI-11 and CITB_978_SKB library.
FEATURES             Location/Qualifiers
     source          1..110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
                     /clone="R-336P14"
                     /clone_lib="RPCI-11"
                     /note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
                     , BAC 563N4 contains Spast gene: exon 4 to exon 17"
     gene            9932..102009
                     /gene="Spast"
     mRNA            9932..102009
                     /gene="Spast"
                     /product="spastin protein"
     exon            9932..10471
                     /gene="Spast"
                     /number=1
     5'UTR           9932..10056
                     /gene="Spast"
     CDS             join(10057..10471,33719..33805,35749..35832,45023..45118,
                     60864..61051,61928..62061,62345..62438,73174..73248,
                     74634..74705,82789..82864,83103..83194,83335..83414,
                     88130..88172,89562..89641,91163..91233,93444..93484,
                     100600..100722)
                     /gene="Spast"
                     /codon_start=1
                     /product="spastin protein"
                     /protein_id="CAB60208.1"
                     /db_xref="GI:6273493"
                     /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                     PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                     SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                     AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                     ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                     GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                     NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                     KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                     REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                     PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                     EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
     intron          10472..33718
                     /gene="Spast"
                     /number=1
     exon            33719..33805
                     /gene="Spast"
                     /number=2
     intron          33806..35748
                     /gene="Spast"
                     /number=2
     exon            35749..35832
                     /gene="Spast"
                     /number=3
     intron          35833..45022
                     /gene="Spast"
                     /number=3
     exon            45023..45118
                     /gene="Spast"
                     /number=4
     intron          45119..60863
                     /gene="Spast"
                     /number=4
     exon            60864..61051
                     /gene="Spast"
                     /number=5
     intron          61052..61927
                     /gene="Spast"
                     /number=5
     exon            61928..62061
                     /gene="Spast"
                     /number=6
     intron          62062..62344
                     /gene="Spast"
                     /number=6
     exon            62345..62438
                     /gene="Spast"
                     /number=7
     intron          62439..73173
                     /gene="Spast"
                     /number=7
     exon            73174..73248
                     /gene="Spast"
                     /number=8
     intron          73249..74633
                     /gene="Spast"
```

```
                     /number=8
     exon            74634..74705
                     /gene="Spast"
                     /number=9
     intron          74706..82788
                     /gene="Spast"
                     /number=9
     exon            82789..82864
                     /gene="Spast"
                     /number=10
     intron          82865..83102
                     /gene="Spast"
                     /number=10
     exon            83103..83194
                     /gene="Spast"
                     /number=11
     intron          83195..83334
                     /gene="Spast"
                     /number=11
     exon            83335..83414
                     /gene="Spast"
                     /number=12
     intron          83415..88129
                     /gene="Spast"
                     /number=12
     exon            88130..88172
                     /gene="Spast"
                     /number=13
     intron          88173..89561
                     /gene="Spast"
                     /number=13
     exon            89562..89641
                     /gene="Spast"
                     /number=14
     intron          89642..91162
                     /gene="Spast"
                     /number=14
     exon            91163..91233
                     /gene="Spast"
                     /number=15
     intron          91234..93443
                     /gene="Spast"
                     /number=15
     exon            93444..93484
                     /gene="Spast"
                     /number=16
     intron          93485..100599
                     /gene="Spast"
                     /number=16
     exon            100600..102009
                     /gene="Spast"
                     /number=17
     3'UTR           100721..102009
                     /gene="Spast"
     polyA_signal    101973..101978
                     /gene="Spast"
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match            100.0%;  Score 501;  DB 9;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 4.8e-113;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db       4000 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 4059

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
Db       4060 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 4119

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 180
Db       4120 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 4179

Qy        181 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 240
Db       4180 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 4239

Qy        241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 300
Db       4240 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 4299

Qy        301 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       4300 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 4359

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 420
Db       4360 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 4419

Qy        421 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 480
Db       4420 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 4479

Qy        481 ATGAATGAGGCTAGCTAGGCC 501
Db       4480 ATGAATGAGGCTAGCTAGGCC 4500


RESULT 4
AC012364
LOCUS       AC012364               155943 bp    DNA     linear   PRI 09-JAN-2002
DEFINITION  Homo sapiens BAC clone RP11-444D15 from 2, complete sequence.
ACCESSION   AC012364
VERSION     AC012364.7  GI:14589736
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155943)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2  (bases 1 to 155943)
  AUTHORS   Ali,J. and Abbott,A.
  TITLE     The sequence of Homo sapiens BAC clone RP11-444D15
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 155943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 155943)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jul 3, 2001 this sequence version replaced gi:11602953.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_NH0444D15
```

```
            ----------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. John D.
            McPherson, Department of Genetics, Washington University, St. Louis
            MO.  For additional information about the map position of this
            sequence, see http://genome.wustl.edu/gsc

            SOURCE INFORMATION:
            The RPCI-11 human BAC library was made from the blood of one male
            donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
            Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
            approach for construction of bacterial artificial chromosome
            libraries.  Genomics 51:1-8.  The clone may be obtained either from
            Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
            and coworkers at the Roswell Park Cancer Institute
            (http://bacpac.med.buffalo.edu)
            VECTOR:  pBACe3.6
            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the left is RP11-431P19; the clone sequenced
            to the right is RP11-78E13.  Actual start of this clone is at base
            position 1 of RP11-444D15; actual end is at base position 155943 of
            RP11-444D15.

            Data from AC011232 and AC010981 was used to finish this clone,
            AC012364.  The sequence fidelity between bases 112626 to 112643 can
            not be guaranteed due to an unresolved homopolymeric run. The
            sequence between 112648 to 112663 is single stranded.
FEATURES             Location/Qualifiers
     source          1. .155943
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-444D15"
                     /clone_lib="RPCI-11"
     repeat_region   310. .617
                     /rpt_family="Alu"
     repeat_region   593. .617
                     /rpt_family="AT_rich"
     repeat_region   728. .748
                     /rpt_family="AT_rich"
     repeat_region   732. .1026
                     /rpt_family="Alu"
     repeat_region   1169. .1240
                     /rpt_family="(TATATG)n"
     repeat_region   1610. .2503
                     /rpt_family="L2"
     repeat_region   2504. .2776
                     /rpt_family="Alu"
     repeat_region   2777. .3498
                     /rpt_family="L2"
     repeat_region   3590. .3742
                     /rpt_family="MIR"
     repeat_region   4452. .4765
                     /rpt_family="Alu"
     misc_feature    4550. .4975
                     /note="match to EST AA632695 (NID:g2556109) np83b02.s1"
     misc_feature    4661. .4692
                     /note="match to EST BG571163 (NID:g13578829)"
     repeat_region   4970. .5239
                     /rpt_family="Alu"
     repeat_region   5211. .5238
                     /rpt_family="(CAAAAA)n"
     misc_feature    5234. .5328
                     /note="match to EST BG571163 (NID:g13578829)"
     misc_feature    5242. .5328
                     /note="match to EST BG757662 (NID:g14068315)"
     misc_feature    5247. .5328
                     /note="match to EST AU121379 (NID:g10936614)"
     misc_feature    5247. .5328
                     /note="match to EST BF205430 (NID:g11099016)"
     misc_feature    5247. .5328
                     /note="match to EST BG033287 (NID:g12425431)"
     misc_feature    5247. .5328
                     /note="similar to Mus musculus EST AA619874 (NID:g2523750)
                     v158g02.s1"
     misc_feature    5247. .5328
                     /note="similar to Mus musculus EST AW120623
                     (NID:g6095956)"
     repeat_region   5740. .6346
                     /rpt_family="L2"
     repeat_region   6347. .6438
                     /rpt_family="AcHobo"
     repeat_region   6453. .6777
                     /rpt_family="L2"
     repeat_region   6938. .7406
                     /rpt_family="L1"
     repeat_region   7415. .8075
                     /rpt_family="ERV1"
     misc_feature    7900. .8421
                     /note="match to EST AW994939 (NID:g8255173)"
     repeat_region   8087. .8245
                     /rpt_family="L1"
     repeat_region   8447. .8524
                     /rpt_family="MER2_type"
     repeat_region   8525. .8799
                     /rpt_family="Alu"
     repeat_region   8830. .9067
                     /rpt_family="MER2_type"
     repeat_region   9917. .10069
                     /rpt_family="MER1_type"
     misc_feature    10293. .10786
                     /note="match to EST N54249 (NID:g1195415) yv65c12.s1"
     repeat_region   10566. .10753
                     /rpt_family="Alu"
     repeat_region   10820. .11873
                     /rpt_family="L1"
     repeat_region   11588. .11607
                     /rpt_family="(A)n"
     repeat_region   11863. .11888
                     /rpt_family="AT_rich"
     repeat_region   11887. .12044
                     /rpt_family="Tc2"
     repeat_region   12045. .12345
                     /rpt_family="Alu"
     repeat_region   12382. .12694
                     /rpt_family="Alu"
     repeat_region   12666. .12727
                     /rpt_family="A-rich"
     repeat_region   12743. .13296
                     /rpt_family="Tc2"
     repeat_region   13463. .13492
                     /rpt_family="(T)n"
     misc_feature    13463. .13478
                     /note="similar to Homo sapiens EST BG110835
                     (NID:g12604341)"
     misc_feature    13820. .14405
                     /note="match to EST BG434159 (NID:g13340653)"
     repeat_region   14067. .14097
                     /rpt_family="AT_rich"
     repeat_region   14793. .15032
```

```
                /rpt_family="L1"
  repeat_region 15033..15330
                /rpt_family="Alu"
  repeat_region 15331..15399
                /rpt_family="L1"
  repeat_region 15400..15700
                /rpt_family="Alu"
  misc_feature  15400..15410
                /note="match to EST BG576844 (NID:g13584497)"
  repeat_region 15701..15745

  Query Match            100.0%;  Score 501;  DB 9;  Length 155943;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-115;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db     119733 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 119792

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
Db     119793 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 119852

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 180
Db     119853 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 119912

Qy        181 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 240
Db     119913 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 119972

Qy        241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 300
Db     119973 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 120032

Qy        301 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db     120033 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 120092

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 420
Db     120093 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 120152

Qy        421 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 480
Db     120153 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 120212

Qy        481 ATGAATGAGGCTAGCTAGGCC 501
Db     120213 ATGAATGAGGCTAGCTAGGCC 120233


RESULT 5
HS106502
LOCUS       HS106502            127192 bp    DNA     linear   PRI 15-MAR-2001
DEFINITION  Human DNA sequence from clone RP5-106502 on chromosome
            20p11.21-11.23 Contains a copy of the RPL41 gene for ribosomal
            protein L41, an ANT2 (fibroblast adenine nucleotide translocator 2
            (ADP,ATP carrier protein, ADP/ATP translocase)) pseudogene, the
            PAX1 gene for paired box gene 1, ESTs, STSs, GSSs, and three CpG
            islands, complete sequence.
ACCESSION   AL035562
VERSION     AL035562.14  GI:5123747
KEYWORDS    HTG; adenine nucleotide translocator; ANT2; CpG island; paired box;
            PAX1; ribosomal protein L41; RPL41.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 127192)
  AUTHORS   Smith,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 22, 1999 this sequence version replaced gi:4808216.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 20, constructed by the Sanger Centre Chromosome 20
            Mapping Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr20
            RP5-106502 is from the library RPCI-5 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP5-106502.
FEATURES             Location/Qualifiers
     source          1..127192
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="RZPD:RPCIP704O021065"
                     /db_xref="taxon:9606"
                     /chromosome="20"
                     /map="p11.21-11.23"
                     /clone="RP5-106502"
                     /clone_lib="RPCI-5"
     misc_feature    2..309
                     /note="match: GSS: Em:AQ224570"
     repeat_region   1083..1406
                     /note="L1PB1 repeat: matches 5822..6152 of consensus"
     misc_feature    complement(1407..1783)
                     /note="match: GSS: Em:AQ424123"
     misc_feature    complement(1714..1746)
                     /note="match: GSS: Em:AQ284374"
     repeat_region   2384..2931
                     /note="MLT1F repeat: matches 12..539 of consensus"
     misc_feature    complement(3228..3649)
                     /note="match: GSS: Em:AQ176778"
     repeat_region   3420..3642
                     /note="MER58A repeat: matches 1..224 of consensus"
     misc_feature    3683..4352
                     /note="match: GSS: Em:AQ474310"
     repeat_region   5113..5157
                     /note="L1ME3 repeat: matches 5746..5791 of consensus"
     repeat_region   5158..5483
                     /note="AluSc repeat: matches 1..304 of consensus"
     repeat_region   5484..5868
                     /note="L1ME3 repeat: matches 5791..6163 of consensus"
     repeat_region   6988..7110
                     /note="L1M4 repeat: matches 5694..5817 of consensus"
     repeat_region   7135..7230
                     /note="MIR repeat: matches 113..211 of consensus"
     repeat_region   7233..7520
                     /note="MLT1F repeat: matches 8..291 of consensus"
     misc_feature    complement(7587..8085)
                     /note="match: GSS: Em:AQ205523"
     misc_feature    8088..8376
                     /note="match: GSS: Em:AQ492709"
     repeat_region   9472..9916
```

```
                     /note="Tigger2a repeat: matches 1. .434 of consensus"
     repeat_region   10755. .10778
                     /note="12 copies 2 mer tt 95% conserved"
     gene            complement(10779. .11191)
                     /gene="dJ106502.1"
     mRNA            complement(10779. .11191)
                     /gene="dJ106502.1"
                     /product="dJ106502.1 (novel protein similar to ribosomal
                     protein L41)"
                     /note="match: cDNAs: Em:Z12962 Em:AF026844 Em:U22229
                     Em:S64030 Em:U93862 Em:X82550 Em:Y11009 Em:D28462
                     match: ESTs: Em:F32536 Em:F25899 Em:F26745 Em:F25237
                     Em:AI123298 Em:AA651967 Em:AA531312 Em:AA634342
                     Em:AA533298 Em:AA534033 Em:AA314794 Em:AA533946
                     Em:AA468914 Em:AA528542 Em:AA531395 Em:AA633421
                     Em:AA652661 Em:AA506196 Em:AA670217 Em:AA484111
                     Em:AA535299 Em:AA484100 Em:AA570180 Em:AA579466
                     Em:AA528396 Em:AI523321 Em:AA316589 Em:W49712 Em:AA650233
                     Em:AA973246 Em:AA973231 Em:AA745030 Em:N30894 Em:AA543019
                     Em:AA316918 Em:AA308031 Em:AA532496 Em:AA316925 Em:H78143
                     Em:AA878405 Em:AA313260 Em:AA377966 Em:AA353795
                     Em:AA338793 Em:AA341533 Em:AA345296 Em:AI009558"
                     /evidence=not_experimental
     misc_feature    10779. .11157
                     /note="match: STS: Em:T60122"
     misc_feature    complement(10779. .10927)
                     /gene="dJ106502.1"
                     /note="match: STS: Em:D11773"
     misc_feature    10779. .10923
                     /note="match: GSS: Em:AQ037469"
     misc_feature    complement(10794. .11165)
                     /gene="dJ106502.1"
                     /note="match: GSS: Em:AF046553"
     polyA_signal    complement(10803. .10808)
                     /gene="dJ106502.1"
     misc_feature    complement(10837. .11191)
                     /gene="dJ106502.1"
                     /note="match: GSS: Em:AQ181846"
     misc_feature    10925. .11101
                     /note="match: GSS: Em:AQ037469"
     misc_feature    complement(10957. .11119)
                     /gene="dJ106502.1"
                     /note="match: STS: Em:N70320"
     misc_feature    11062. .11190
                     /note="match: GSS: Em:B16341 Em:B73682"
     misc_feature    11071. .11190
                     /note="match: GSS: Em:B17936"
     CDS             complement(11074. .11151)
                     /gene="dJ106502.1"
                     /note="translation of ORF identical to original RPL41 gene
                     located on chromosome 12, could be a pseudogene"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="dJ106502.1 (novel protein similar to ribosomal
                     protein L41)"
                     /protein_id="CAB46994.1"
                     /db_xref="GI:5457305"
                     /db_xref="REMTREMBL:CAB46994"
                     /translation="MRAKWRKKRMRRLKRKRRKMRQRSK"
     repeat_region   11314. .11404
                     /note="MIR repeat: matches 47. .147 of consensus"
     repeat_region   11434. .11723
                     /note="AluJb repeat: matches 1. .282 of consensus"
     repeat_region   12266. .12824
                     /note="L1M3c repeat: matches 980. .1595 of consensus"
     repeat_region   12926. .13551
                     /note="L1M4 repeat: matches 2570. .3208 of consensus"
     repeat_region   13684. .13840
                     /note="Tigger3b repeat: matches 238. .376 of consensus"
     repeat_region   13924. .14026
                     /note="Tigger3(Golem) repeat: matches 1. .100 of consensus"
     repeat_region   14048. .14518
                     /note="L1 repeat: matches 3623. .4097 of consensus"
     repeat_region   14698. .14867
                     /note="L1M4 repeat: matches 4332. .4501 of consensus"
     repeat_region   14869. .15328
                     /note="Tigger3b repeat: matches 2. .486 of consensus"
     repeat_region   15330. .20729
                     /note="L1PA6 repeat: matches 753. .6140 of consensus"
     repeat_region   20732. .21121
                     /note="Tigger3b repeat: matches 584. .930 of consensus"
     repeat_region   21122. .21540
                     /note="MSTA repeat: matches 1. .426 of consensus"
     repeat_region   21541. .21828
                     /note="Tigger3b repeat: matches 930. .1230 of consensus"
     repeat_region   21829. .22049
                     /note="L1M4 repeat: matches 4499. .4736 of consensus"
     repeat_region   22076. .22378
                     /note="AluY repeat: matches 3. .304 of consensus"
     repeat_region   22428. .22753
                     /note="L1MEc repeat: matches 1419. .1737 of consensus"
     repeat_region   22754. .23086
                     /note="THE1B repeat: matches 1. .364 of consensus"
     repeat_region   23087. .23224
                     /note="L1MEc repeat: matches 1737. .1876 of consensus"
     gene            23344. .24068
                     /gene="dJ106502.2"
                     /pseudo
     CDS             23344. .24068
                     /gene="dJ106502.2"
                     /note="dJ106502.2 (ANT2 (adenine nucleotide translocator 2
                     (fibroblast) (ADP,ATP carrier protein, ADP/ATP
                     translocase)) pseudogene)
                     match: cDNAs: Em:J03592 Em:M24103 Em:D12771 Em:J02683
                     Em:X74510 Em:D12770 Em:U27315 Em:M24102 Em:J02966
                     Em:AB008456 Em:AB008462 Em:AF091236
                     match: ESTs: Em:AA209443 Em:AA143431 Em:R25296 Em:R13538
                     Em:AA088746 Em:W03111 Em:R72717 Em:AA204798 Em:W02616

  Query Match             19.7%;  Score 98.8;  DB 9;  Length 127192;
  Best Local Similarity   70.0%;  Pred. No. 2.9e-14;
  Matches  147;  Conservative    0;  Mismatches   62;  Indels    1;  Gaps    1;

Qy        189 CAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCT 248
              ||    || :|||:  ||   .|| :  ||||`:||| `: :||||.: |||     ||
Db      41215 CATAAGTTTCTCCCATCCAGGTATGCATGCCCCTTTGTAATGTGACTTTGCTGTTCTTCC 41274

Qy        249 ATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTT 308
              |',||'  || ::·  |` |||  `:| | |`| |     ||||`::    ||| .||||
Db      41275 ATAAAGAGGTAGAGTCTACTTTATCATCTCTTGAATCTGGGTTGGCCTGGTGATTTGCTT 41334

Qy        309 TGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGA 367
                ||;  | |.`||  : | |: :||  ` ` | ||:·:·||  !   : | | :  ||||'
Db      41335 TGACCATTAGAATGCTGTAGAAGTGATATTATCCAACTTTCAAGGCTGGCTCTTAAGAGA 41394

Qy        368 ACTTACACCTTCCACTCACACTCTCTTGGA 397
               ||· |  . |||  |  ::  |||`  ||
Db      41395 CCTTGCGGCTTCCACTGTCACCCTCTTAGA 41424


RESULT 6
AC091063
LOCUS       AC091063               73096 bp    DNA     linear   HTG 25-MAR-2001
DEFINITION  Homo sapiens chromosome 18 clone RP11-1078M23 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC091063
VERSION     AC091063.1  GI:13446315
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 73096)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-1078M23
  JOURNAL   Unpublished
```

REFERENCE 2 (bases 1 to 73096)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Yo,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-------------- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
Center project name: I12746
Center clone name: 1078_M_23
--------------.
* NOTE: This record contains 88 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 721: contig of 721 bp in length
* 722 821: gap of 100 bp
* 822 1559: contig of 738 bp in length
* 1560 1659: gap of 100 bp
* 1660 2369: contig of 710 bp in length
* 2370 2469: gap of 100 bp
* 2470 3207: contig of 738 bp in length
* 3208 3307: gap of 100 bp
* 3308 4097: contig of 790 bp in length
* 4098 4197: gap of 100 bp
* 4198 4925: contig of 728 bp in length
* 4926 5025: gap of 100 bp
* 5026 5765: contig of 740 bp in length
* 5766 5865: gap of 100 bp
* 5966 6409: contig of 544 bp in length
* 6410 6509: gap of 100 bp
* 6510 7252: contig of 743 bp in length
* 7253 7352: gap of 100 bp
* 7353 8096: contig of 744 bp in length
* 8097 8196: gap of 100 bp
* 8197 8935: contig of 739 bp in length
* 8936 9035: gap of 100 bp
* 9036 9784: contig of 749 bp in length
* 9785 9884: gap of 100 bp
* 9885 10623: contig of 739 bp in length
* 10624 10723: gap of 100 bp
* 10724 11511: contig of 788 bp in length
* 11512 11611: gap of 100 bp
* 11612 12322: contig of 711 bp in length
* 12323 12422: gap of 100 bp
* 12423 13141: contig of 719 bp in length
* 13142 13241: gap of 100 bp
* 13242 13940: contig of 699 bp in length
* 13941 14040: gap of 100 bp
* 14041 14765: contig of 725 bp in length
* 14766 14865: gap of 100 bp
* 14866 15599: contig of 734 bp in length
* 15600 15699: gap of 100 bp
* 15700 16387: contig of 688 bp in length
* 16388 16487: gap of 100 bp
* 16488 17217: contig of 730 bp in length
* 17218 17317: gap of 100 bp
* 17318 18072: contig of 755 bp in length
* 18073 18172: gap of 100 bp
* 18173 18914: contig of 742 bp in length
* 18915 19014: gap of 100 bp
* 19015 19739: contig of 725 bp in length
* 19740 19839: gap of 100 bp
* 19840 20627: contig of 788 bp in length
* 20628 20727: gap of 100 bp
* 20728 21466: contig of 739 bp in length
* 21467 21566: gap of 100 bp
* 21567 22295: contig of 729 bp in length
* 22296 22395: gap of 100 bp
* 22396 23128: contig of 733 bp in length
* 23129 23228: gap of 100 bp
* 23229 23959: contig of 731 bp in length
* 23960 24059: gap of 100 bp
* 24060 24772: contig of 713 bp in length
* 24773 24872: gap of 100 bp
* 24873 25617: contig of 745 bp in length
* 25618 25717: gap of 100 bp
* 25718 26460: contig of 743 bp in length
* 26461 26560: gap of 100 bp
* 26561 27286: contig of 726 bp in length
* 27287 27386: gap of 100 bp
* 27387 28120: contig of 734 bp in length
* 28121 28220: gap of 100 bp
* 28221 28952: contig of 732 bp in length
* 28953 29052: gap of 100 bp
* 29053 29802: contig of 750 bp in length
* 29803 29902: gap of 100 bp
* 29903 30643: contig of 741 bp in length
* 30644 30743: gap of 100 bp
* 30744 31471: contig of 728 bp in length
* 31472 31571: gap of 100 bp
* 31572 32298: contig of 727 bp in length
* 32299 32398: gap of 100 bp
* 32399 33125: contig of 727 bp in length
* 33126 33225: gap of 100 bp
* 33226 33959: contig of 734 bp in length
* 33960 34059: gap of 100 bp
* 34060 34762: contig of 703 bp in length
* 34763 34862: gap of 100 bp
* 34863 35587: contig of 725 bp in length
* 35588 35687: gap of 100 bp
* 35688 36416: contig of 729 bp in length
* 36417 36516: gap of 100 bp
* 36517 37255: contig of 739 bp in length
* 37256 37355: gap of 100 bp
* 37356 38096: contig of 741 bp in length
* 38097 38196: gap of 100 bp
* 38197 38947: contig of 751 bp in length
* 38948 39047: gap of 100 bp
* 39048 39785: contig of 738 bp in length
* 39786 39885: gap of 100 bp
* 39886 40626: contig of 741 bp in length
* 40627 40726: gap of 100 bp

```
     •    40727    41461: contig of 735 bp in length
     •    41462    41561: gap of 100 bp
     •    41562    42299: contig of 738 bp in length
     •    42300    42399: gap of 100 bp
     •    42400    43144: contig of 745 bp in length
     •    43145    43244: gap of 100 bp
     •    43245    43990: contig of 746 bp in length
     •    43991    44090: gap of 100 bp
     •    44091    44811: contig of 721 bp in length
     •    44812    44911: gap of 100 bp
     •    44912    45616: contig of 705 bp in length
     •    45617    45716: gap of 100 bp
     •    45717    46450: contig of 734 bp in length
     •    46451    46550: gap of 100 bp
     •    46551    47298: contig of 748 bp in length
     •    47299    47398: gap of 100 bp
     •    47399    48145: contig of 747 bp in length
     •    48146    48245: gap of 100 bp
     •    48246    48959: contig of 714 bp in length
     •    48960    49059: gap of 100 bp
     •    49060    49786: contig of 727 bp in length
     •    49787    49886: gap of 100 bp
     •    49887    50602: contig of 716 bp in length
     •    50603    50702: gap of 100 bp
     •    50703    51337: contig of 635 bp in length
     •    51338    51437: gap of 100 bp
     •    51438    52176: contig of 739 bp in length
     •    52177    52276: gap of 100 bp
     •    52277    53020: contig of 744 bp in length
     •    53021    53120: gap of 100 bp
     •    53121    53855: contig of 735 bp in length
     •    53856    53955: gap of 100 bp
     •    53956    54695: contig of 740 bp in length
     •    54696    54795: gap of 100 bp
     •    54796    55539: contig of 744 bp in length
     •    55540    55639: gap of 100 bp
     •    55640    56363: contig of 724 bp in length
     •    56364    56463: gap of 100 bp
     •    56464    57216: contig of 753 bp in length
     •    57217    57316: gap of 100 bp

  Query Match             19.4%;  Score 97.4;  DB 2;  Length 73096;
  Best Local Similarity   70.0%;  Pred. No. 7.2e-14;
  Matches  145;  Conservative    0;  Mismatches   61;  Indels    1;  Gaps    1;

Qy        189 CAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCT 248
              ||  | |  ||||  | ||      ||       ||||||||||  ||| ||||  |||  |
Db      46185 CATAAGTTTCTCCCATCCAGGTATGCATGCCCCTTTGTAATGTGACTTTGCTGTTCTTCC 46244

Qy        249 ATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTT 308
              |  ||| |  ||    |||   ||  ||    |  | |||| || ||   ||||
Db      46245 ATAAAGAGGTAGAGTCTACTTTATCATCTCTTGAATCTGGGTTGGCCTGGTGATTTGCTT 46304

Qy        309 TGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGA 367
              |||  |  |||   |||   | |    |   ||  ||          | |     |  ||
Db      46305 TGACCATTAGAATGCTGTAGAAGTGATATTATCCAACTTTCAAGGCTGGCTCTTAAGAGA 46364

Qy        368 ACTTACACCTTCCACTCACACTCTCTT 394
                 |  |   |||||||  ||   ||||
Db      46365 CCTTGCGGCTTCCACTGTCACCCTCTT 46391


RESULT 7
AL158158/c
LOCUS       AL158158              194835 bp    DNA     linear   PRI 05-APR-2001
DEFINITION  Human DNA sequence from clone RP11-427L11 on chromosome 9q31.2-32,
            complete sequence.
ACCESSION   AL158158
VERSION     AL158158.14  GI:13559997
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194835)
  AUTHORS   Williams,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-APR-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Apr 6, 2001 this sequence version replaced gi:12733508.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-427L11 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            This sequence is the entire insert of clone RP11-427L11.
FEATURES             Location/Qualifiers
     source          1..194835
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /map="q31.2-32"
                     /clone="RP11-427L11"
                     /clone_lib="RPCI-11.2"
     repeat_region   1..224
                     /note="L1P repeat: matches 3..226 of consensus"
     repeat_region   326..615
                     /note="AluSg repeat: matches 1..299 of consensus"
     repeat_region   627..772
                     /note="AluSg/x repeat: matches 158..301 of consensus"
     repeat_region   929..1168
                     /note="L1M4 repeat: matches 5149..5403 of consensus"
     repeat_region   1195..1519
                     /note="AluSx repeat: matches 1..311 of consensus"
     repeat_region   1698..1726
                     /note="MER5A repeat: matches 16..44 of consensus"
     repeat_region   1728..1868
                     /note="AluSg/x repeat: matches 150..302 of consensus"
     repeat_region   1887..1932
                     /note="MER5A repeat: matches 61..106 of consensus"
     repeat_region   1966..2242
                     /note="L1ME3A repeat: matches 5614..5894 of consensus"
     misc_feature    complement(1992..2500)
                     /note="match: GSS: Em:AQ172933"
     repeat_region   2308..2619
                     /note="AluSp repeat: matches 1..313 of consensus"
     repeat_region   2707..2822
                     /note="L1ME3A repeat: matches 6017..6137 of consensus"
     repeat_region   3336..3403
                     /note="34 copies 2 mer cc 66% conserved"
     repeat_region   3420..3725
                     /note="AluSx repeat: matches 1..300 of consensus"
     repeat_region   4319..4467
```

```
                     /note="MER5B repeat: matches 4..175 of consensus"
     repeat_region   4884..5309
                     /note="L2 repeat: matches 166..616 of consensus"
     repeat_region   5596..5870
                     /note="L1MA9 repeat: matches 6027..6305 of consensus"
     repeat_region   5871..5938
                     /note="AluJ/FLAM repeat: matches 1..68 of consensus"
     repeat_region   5967..6648
                     /note="L2 repeat: matches 1284..2090 of consensus"
     repeat_region   6649..6948
                     /note="AluSq repeat: matches 1..300 of consensus"
     repeat_region   6949..7003
                     /note="L2 repeat: matches 2090..2136 of consensus"
     repeat_region   7072..7462
                     /note="L2 repeat: matches 2295..2695 of consensus"
     repeat_region   7588..7789
                     /note="MIR repeat: matches 12..236 of consensus"
     misc_feature    complement(8285..8469)
                     /note="match: STS: Em:G35863"
     misc_feature    complement(8329..8496)
                     /note="match: GSS: Em:AQ227387"
     misc_feature    complement(8329..8495)
                     /note="match: GSS: Em:B65750"
     misc_feature    complement(8339..8502)
                     /note="match: GSS: Em:AQ110861"
     misc_feature    8354..8529
                     /note="match: STS: Em:G54038"
     misc_feature    join(8370..8532,9031..9156,15133..15132,15671..15768)
                     /note="match: STS: Em:G29920"
     repeat_region   8591..8662
                     /note="36 copies 2 mer ta 81% conserved"
     repeat_region   9732..9893
                     /note="AluJ repeat: matches 130..287 of consensus"
     misc_feature    complement(9737..9894)
                     /note="match: GSS: Em:AQ281101"
     misc_feature    complement(9886..10119)
                     /note="match: GSS: Em:B53125"
     misc_feature    complement(9897..10091)
                     /note="match: STS: Em:HSA108WE1"
     misc_feature    complement(9907..10123)
                     /note="match: GSS: Em:B17453"
     repeat_region   9932..9967
                     /note="18 copies 2 mer ta 100% conserved"
     repeat_region   9973..10105
                     /note="AluJ repeat: matches 1..133 of consensus"
     repeat_region   10133..10392
                     /note="AluSx repeat: matches 38..299 of consensus"
     misc_feature    10183..10319
                     /note="match: GSS: Em:AQ782105"
     repeat_region   10404..10492
                     /note="MER5B repeat: matches 7..97 of consensus"
     misc_feature    10992..11078
                     /note="match: GSS: Em:AQ544271"
     repeat_region   11354..11653
                     /note="AluSx repeat: matches 1..300 of consensus"
     repeat_region   11742..12929
                     /note="Tigger3b repeat: matches 2..1231 of consensus"
     repeat_region   12987..13146
                     /note="MIR repeat: matches 89..248 of consensus"
     repeat_region   13153..13468
                     /note="AluSx repeat: matches 1..312 of consensus"
     repeat_region   13678..13848
                     /note="L2 repeat: matches 2537..2703 of consensus"
     repeat_region   14394..14609
                     /note="MIR repeat: matches 12..257 of consensus"
     misc_feature    15816..15902
                     /note="match: GSS: Em:B46502"
     repeat_region   15900..15947
                     /note="24 copies 2 mer aa 79% conserved"
     repeat_region   16257..16558
                     /note="AluSg repeat: matches 1..302 of consensus"
     repeat_region   16956..17220
                     /note="AluSq repeat: matches 38..304 of consensus"
     repeat_region   17415..17489
                     /note="37 copies 2 mer aa 71% conserved"
     repeat_region   18552..18845
                     /note="AluSq repeat: matches 1..293 of consensus"
     repeat_region   19629..19689
                     /note="L2 repeat: matches 2685..2743 of consensus"
     misc_feature    19779..21330
                     /note="CpG island"
                     /evidence=not_experimental
     repeat_region   21481..21679
                     /note="MIR repeat: matches 48..262 of consensus"
     repeat_region   22591..22818
                     /note="MIR repeat: matches 23..258 of consensus"
     misc_feature    complement(23067..23458)
                     /note="match: GSS: Em:B43385"
     misc_feature    complement(23074..23458)
                     /note="match: GSS: Em:B43132"
     repeat_region   23731..23893
                     /note="MER5A repeat: matches 17..180 of consensus"
     misc_feature    24008..24365
                     /note="match: GSS: Em:AQ088039"
     repeat_region   24107..24196
                     /note="MIR repeat: matches 105..202 of consensus"
     repeat_region   24953..25091
                     /note="L1MA8 repeat: matches 6145..6287 of consensus"
     repeat_region   25184..25589
                     /note="MER50 repeat: matches 1..436 of consensus"
     repeat_region   25620..25718
                     /note="MER50 repeat: matches 640..734 of consensus"
     repeat_region   25839..25991
                     /note="L1ME1 repeat: matches 5212..5366 of consensus"
     repeat_region   25992..26078
                     /note="AluSp/q repeat: matches 174..260 of consensus"
     repeat_region   26084..26798
                     /note="L1ME1 repeat: matches 5347..6036 of consensus"
     repeat_region   27326..27430
                     /note="L1MB6 repeat: matches 4066..4176 of consensus"
     repeat_region   27431..27740
                     /note="AluSx repeat: matches 1..310 of consensus"
     repeat_region   27741..28086
                     /note="L1MB6 repeat: matches 4176..4455 of consensus"
     repeat_region   28087..28226
                     /note="AluJb repeat: matches 3..132 of consensus"
     repeat_region   28227..28530
                     /note="AluSx repeat: matches 1..297 of consensus"

  Query Match             18.8%;  Score 94;  DB 9;  Length 194835;
  Best Local Similarity   62.4%;  Pred. No. 4.1e-13;
  Matches  181;  Conservative    0;  Mismatches  105;  Indels    4;  Gaps    2;

Qy        151 ACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCAT 210

Db      33983 AAATAAGGACTAGAGTAGATTGCTTACAAAGATGGTCCCAGTGATTCATCTAATATTTAC 33924

Qy        211 ATATATGCCCCTTTGCAATGTGACTTTGCTACTTC-TCTATCAAGATGTGGAGCTTATTT 269

Db      33923 AGGTCAGCACCTTTGCAATATAACTGTGATACTTCTTGAATAAAGAGATGTAGTCTATTT 33864

Qy        270 TCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAA 329

Db      33863 CCCCATCCCTTGAAACTGGATAGGCCTTGTGACTTTTTCTGACCAGTAGGTACAGCAGAA 33804

Qy        330 ATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACT 389

Db      33803 GTAATATTGTAGGACTTACAAATCTAGACCCCAAGAAAGCTTGTAGCTTCCATTCTCACC 33744

Qy        390 CTCTTGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAG 439

Db      33743 CTCTTCAAACAC---TGCCCTGAAAAGAAGCTCAGTCTAACCTACTGGAG 33697


RESULT 8
HSDJ792G4/c
```

```
LOCUS       HSDJ792G4             142728 bp    DNA     linear   PRI 31-DEC 2007
DEFINITION  Human DNA sequence from clone RP4-792G4 on chromosome 1p31.2-32.2
            Contains a transcription factor gene, a pseudogene similar to
            tetratricopeptide protein, ESTs, CA repeat (D18438), STSs, GSSs and
            CpG Islands, complete sequence.
ACCESSION   AL049636
VERSION     AL049636.22  GI:11967799
KEYWORDS    HTG; D18438.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142728)
  AUTHORS   Howden,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 22, 2000 this sequence version replaced gi:5531522.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements.  Where the sequence is
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr1
            RP4-792G4 is from the library RPCI-4 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP4-792G4.
FEATURES             Location/Qualifiers
     source          1..142728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /map="p31.2-32.2"
                     /clone="RP4-792G4"
                     /clone_lib="RPCI-4"
     repeat_region   1..62
                     /note="Alu repeat: matches 1..62 of consensus"
     repeat_region   1724..1797
                     /note="Charlie4 repeat: matches 1883..1955 of consensus"
     repeat_region   1799..1842
                     /note="22 copies 2 mer at 84% conserved"
     repeat_region   2387..2414
                     /note="14 copies 2 mer tc 89% conserved"
     misc_feature    10056..10503
                     /note="match: GSS: Em:AQ165153"
     repeat_region   13213..13497
                     /note="AluSg1 repeat: matches 1..309 of consensus"
     repeat_region   13619..13930
                     /note="AluYa8 repeat: matches 1..308 of consensus"
     repeat_region   14743..17426
                     /note="L1M1 repeat: matches -879..1556 of consensus"
     repeat_region   17806..17891
                     /note="L1M1 repeat: matches 1540..1625 of consensus"
     repeat_region   18385..19497
                     /note="L1M1 repeat: matches 1623..2761 of consensus"
     misc_feature    23778..24163
                     /note="match: GSS: Em:AQ270189"
     repeat_region   24773..24820
                     /note="24 copies 2 mer at 83% conserved"
     repeat_region   25736..25769
                     /note="17 copies 2 mer ta 82% conserved"
     repeat_region   27577..27630
                     /note="27 copies 2 mer ta 74% conserved"
     repeat_region   27634..27880
                     /note="MER2 repeat: matches 14..255 of consensus"
     repeat_region   28476..28606
                     /note="AluSq/x repeat: matches 3..133 of consensus"
     repeat_region   31013..31234
                     /note="L1MC1 repeat: matches 5946..6172 of consensus"
     repeat_region   31235..31323
                     /note="MER44C repeat: matches 1..89 of consensus"
     repeat_region   31322..31503
                     /note="MER44C repeat: matches 235..417 of consensus"
     repeat_region   31504..31670
                     /note="FRAM repeat: matches 1..168 of consensus"
     repeat_region   31674..31962
                     /note="MER44C repeat: matches 413..697 of consensus"
     repeat_region   31977..32132
                     /note="L1MC1 repeat: matches 6169..6323 of consensus"
     repeat_region   32644..32930
                     /note="AluYa5 repeat: matches 1..301 of consensus"
     repeat_region   35861..35997
                     /note="AluSp/q repeat: matches 175..311 of consensus"
     repeat_region   36422..36492
                     /note="MLT1-INTERNAL repeat: matches 808..861 of
                     consensus"
     repeat_region   36772..36859
                     /note="MLT1-INTERNAL repeat: matches 1083..1157 of
                     consensus"
     repeat_region   40721..40839
                     /note="MER33 repeat: matches 1..113 of consensus"
     repeat_region   41156..41263
                     /note="MER33 repeat: matches 113..238 of consensus"
     repeat_region   41326..41485
                     /note="L1MA2 repeat: matches 6151..6304 of consensus"
     repeat_region   41795..41903
                     /note="L1MA2 repeat: matches 6041..6151 of consensus"
     repeat_region   41904..41957
                     /note="MER33 repeat: matches 243..297 of consensus"
     repeat_region   42270..42307
                     /note="19 copies 2 mer aa 81% conserved"
     misc_feature    44966..45522
                     /note="match: GSS: Em:AQ422993"
     misc_feature    46018..64094
                     /note="CpG island"
                     /evidence=not_experimental
     misc_feature    complement(46601..47101)
                     /note="match: GSS: Em:AZ016581"
     gene            52353..53789
                     /gene="dJ792G4.1"
     CDS             52353..53789
                     /gene="dJ792G4.1"
                     /note="match: cDNAs: Em:AF052249 Em:U39840 Em:U13220
                     Em:U59832 Em:AF071554 Em:U04198 Em:X01460 Em:AF052251
                     Em:U11690 Em:AF104902 Em:U37272 Em:U87393 Em:X92592
                     Em:M95603 Em:L38607 Em:L13202 Em:L13201 Em:X99291
                     Em:L22760 Em:AB001572 Em:M93128 Em:S79041 Em:AJ011652
                     Em:AF023915 Em:X74143 Em:U90538 Em:X55955 Em:S82462
                     Em:AF072889 Em:L12142 Em:L12143 Em:M14511 Em:L12703
                     Em:U13222 Em:AB028021
                     match: ESTs: Em:AI739311 Em:AI066718 Em:AI554075
                     Em:AA048965 Em:AI391745 Em:AI323408
                     match: proteins: Sw:P79772 Sw:Q61060 Sw:Q16676 Sw:P11087
                     Sw:Q61345 Sw:Q64732 Sw:P55318 Sw:P02452 Tr:O93613
                     Sw:P02453 Sw:Q01149 Tr:Q51415 Sw:P16791 Sw:P26372
                     Tr:Q92751 Tr:O73784 Sw:Q99958 Sw:Q98937 Sw:Q12950
                     Sw:Q63244 Tr:O70220 Sw:P06924 Sw:Q63245 Sw:Q64733
                     Tr:O88778 Tr:O73779 Wp:CE07013 Sw:P55787 Sw:Q60688
```

```
                  Wp:CE02166 Tr:O88737 Sw:O35876 Sw:P03186 Sw:Q63251
                  Tr:O73781 Tr:Q63729 Sw:O00358 Sw:P48634 Tr:O73785
                  Sw:P81489 Tr:O95886 Tr:O35053 Sw:P08121 Tr:O93312
                  Sw:P02461 Sw:P32028 Sw:P20849 Tr:O46046 Tr:O42350
                  Tr:Q26052 Tr:O93440 Sw:Q03391 Sw:P04258 Sw:Q61850
                  Sw:Q61572 Tr:O93529"
                  /codon_start=1
                  /evidence=not_experimental
                  /product="dJ792G4.1 (TRANSCRIPTION FACTOR)"
                  /protein_id="CAC19431.1"
                  /db_xref="GI:11967800"
                  /translation="MTLSGGGSASDMSGQTVLTAEDVDIDVVGEGDDGLEEKDSDAGC
                  DSPAGPPELRLDEADEVPPAAPHHGQPQPPHQQPLTLPKEAAGAGAGPGGDVGAPEAD
                  GCKGGVGGEEGGASGGGPGAGSGSAGGLAPSKPKNSLVKPPYSYIALITMAILQSPQK
                  KLTLSGICEFISNRFPYYREKFPAWQNSIRHNLSLNDCFVKIPREPGNPGKGNYWTLD
                  PQSEDMFDNGSFLRRRKRFKRHQQEHLREQTALMMQSFGAYSLAAAAGAAGPYGRPYG
                  LHPAAAAGAYSHPAAAAAAAAAAALQYPYALPPVAPVLPPAVPLLPSGELGRKAAAFG
                  SQLGPGLQLQLNSLGAAAAAAGTAGAAGTTASLIKSEPSARPSFSIENIIGGGPAAPG
                  GSAVGAGVAGGTGGSGGGSTAQSFLRPPGTVQSAALMATHQPLSLSRTTATIAPILSV
                  PLSGQFLQPAASAAAAAAAAAAQAKWPAQ"
     repeat_region 52601..52739
                  /note="69 copies 2 mer gg 58% conserved"
     misc_feature 56106..56580
                  /note="match: GSS: Em:AQ785165"
     misc_feature 58987..59763
                  /note="CpG island"
                  /evidence=not_experimental
     repeat_region 60970..61005
                  /note="18 copies 2 mer tt 80% conserved"
     misc_feature complement(62750..63178)
                  /note="match: GSS: Em:AQ150149"
     repeat_region 63054..63110
                  /note="U7 repeat: matches 2..59 of consensus"
     misc_feature complement(63549..63891)
                  /note="match: STS: Em:L16339"
     repeat_region 63696..63821
                  /note="63 copies 2 mer tt 70% conserved"
     repeat_region 64593..64646
                  /note="27 copies 2 mer tg 96% conserved"
     repeat_region 67253..67294
                  /note="HY3 repeat: matches 1..41 of consensus"
     repeat_region 67609..67664
                  /note="HY3 repeat: matches 41..100 of consensus"
     repeat_region 69418..69459
                  /note="Alu repeat: matches 251..255 of consensus"
     repeat_region 69738..69786
                  /note="Alu repeat: matches 255..301 of consensus"
     repeat_region 71195..71220
                  /note="13 copies 2 mer tt 100% conserved"
     misc_feature complement(71415..71797)
                  /note="match: GSS: Em:AQ797602"
     repeat_region 71702..71737
                  /note="MER81 repeat: matches 72..114 of consensus"
     repeat_region 73309..73390

  Query Match             18.7%;  Score 93.6;  DB 9;  Length 142728;
  Best Local Similarity   65.9%;  Pred. No. 5.5e-13;
  Matches  166;  Conservative    0;  Mismatches   84;  Indels    2;  Gaps    2;

Qy        246 TCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTG 305
              |   |||  ||  || |||  ||   ||      ||  |  ||   ||  ||  |||  |
Db      80022 TCCAGTAAGAAGTGCAGTTTATTTTGCCAATCCTTGAATCAGGGCTGGCCTTGTGACTTT 79963

Qy        306 CTTTGAC-AATGGAATGTAGTACAAATGACAC-TGTGCAACTTTGGATTTTAGGTTTCGA 363
               |||||  |  ||||  | | | ||       | ||  |  ||  | | |  |
Db      79962 CTTTGACTAATAGAATGTGGCAGAAGTGTTGTATGAGCTCCTCTGGAGTCTAGGCCTCAG 79903

Qy        364 GAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGTCAG 423
              | | | || || | |  |   ||||||  |||  | |         || ||   |
Db      79902 GAAACCTTGCAGCGTCTGCTCTTGCTCTCTTTGAACACTGGCATCGTACAAAGAAGCCCA 79843

Qy        424 GGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACATATG 483
              || || ||||   ||||| |||| |||   |  ||||  |||  |  ||      || |
Db      79842 GGTTAACCTGTAGGAGACAGATGTCTCAGGTGACAGCCAGCATCAACAGCTAGACATTTA 79783

Qy        484 AATGAGGCTAGC 495
              |   |||  | |
Db      79782 AGTGAGGTCATC 79771


RESULT 9
BX248112
LOCUS       BX248112              75046 bp    DNA     linear   PRI 20-FEB-2003
DEFINITION  Human DNA sequence from clone CTD-2340F8 on chromosome 9, complete
            sequence.
ACCESSION   BX248112
VERSION     BX248112.4  GI:28460379
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75046)
  AUTHORS   Sehra,H.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-FEB-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 21, 2003 this sequence version replaced gi:28300993.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            CTD-2340F8 is from the CalTech genomic sperm BAC library D. VECTOR:
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..75046
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="CTD-2340F8"
                     /clone_lib="CIT-HSP-D1"
BASE COUNT    22718 a  14985 c  14997 g  22346 t
ORIGIN

  Query Match             17.0%;  Score 85;  DB 9;  Length 75046;
  Best Local Similarity   62.5%;  Pred. No. 7.9e-11;
  Matches  168;  Conservative    0;  Mismatches   95;  Indels    6;  Gaps    2;

Qy        144 ATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTA 203
```

```
Db      18187 AATTTGAAGATGTCCTATACGGTCGATTGTTTCCAAAGACGACTGCAACACTCCATTCTA 18246

Qy        204 TCCTCATATATATGCCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGC 263

Db      18247 TC-----TTACATGCCCTTTTGCAATATGACTTTTCCACTTTTCCATCAAGACATAGACT 18301

Qy        264 TTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGT 322

Db      18302 CTATCTTACCTCCCCTTAAATCTGGACTAATCTTATGAGTTGCTTTAACCAATATCATGT 18361

Qy        323 AGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCAC 382

Db      18362 GGTAGAAATGACAATGTGCAACTCCAAAGACAAAACTTCAAGAATCCCCACACCTTCCAT 18421

Qy        383 TCACACTCTCTTGGAAACCAGATGCAATG 411

Db      18422 TTTTTATGTCTTGGAATGCAGCTGCCATG 18450


RESULT 10
AL953911
LOCUS       AL953911              150770 bp    DNA     linear   PRI 13-DEC-2002
DEFINITION  Human DNA sequence from clone RP11-242D12 on chromosome 9, complete
            sequence.
ACCESSION   AL953911
VERSION     AL953911.4  GI:26788317
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 150770)
  AUTHORS   Skuce,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 13, 2002 this sequence version replaced gi:25137059.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr9
            RP11-242D12 is from the library RPCI-11.1 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..150770
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-242D12"
                     /clone_lib="RPCI-11.1"
BASE COUNT    48356 a  30391 c  30483 g  44540 t
ORIGIN

  Query Match           17.0%;  Score 85;  DB 9;  Length 150770;
  Best Local Similarity   62.5%;  Pred. No. 7e-11;
  Matches  168;  Conservative    0;  Mismatches   95;  Indels    6;  Gaps    2;

Qy        144 ATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTA 203

Db      96018 AATTTGAAGATGTCCTATACGGTCGATTGTTTCCAAAGACGACTGCAACACTCCATTCTA 96077

Qy        204 TCCTCATATATATGCCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGC 263

Db      96078 TC-----TTACATGCCCTTTTGCAATATGACTTTTCCACTTTTCCATCAAGACATAGACT 96132

Qy        264 TTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGT 322

Db      96133 CTATCTTACCTCCCCTTAAATCTGGACTAATCTTATGAGTTGCTTTAACCAATATCATGT 96192

Qy        323 AGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCAC 382

Db      96193 GGTAGAAATGACAATGTGCAACTCCAAAGACAAAACTTCAAGAATCCCCACACCTTCCAT 96252

Qy        383 TCACACTCTCTTGGAAACCAGATGCAATG 411

Db      96253 TTTTTATGTCTTGGAATGCAGCTGCCATG 96281


RESULT 11
BX088588/c
LOCUS       BX088588              187035 bp    DNA     linear   HTG 08-APR-2003
DEFINITION  Homo sapiens chromosome 9 clone RP11-50I9, *** SEQUENCING IN
            PROGRESS ***.
ACCESSION   BX088588
VERSION     BX088588.4  GI:27805046
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 187035)
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 20, 2003 this sequence version replaced gi:27803285.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA50I9
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 187023 bases at least Q40
            Consensus quality: 187031 bases at least Q30
            Consensus quality: 187034 bases at least Q20
            Insert size: 187035; sum-of-contigs
            Insert size: 196616; 2.8% error; agarose-fp
            Quality coverage: 16.59x in Q20 bases; sum-of-contigs Quality
            coverage: 15.87x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
```

```
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1  187035: contig of 187035 bp in length.
FEATURES             Location/Qualifiers
     source          1..187035
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="9"
                     /clone="RP11-50I9"
                     /clone_lib="RPCI-11.1"
     misc_feature    1..187035
                     /note="assembly_fragment:01460"
BASE COUNT    55353 a  37202 c  37514 g  56966 t
ORIGIN

  Query Match             17.0%;  Score 85;  DB 2;  Length 187035;
  Best Local Similarity   62.5%;  Pred. No. 6.8e-11;
  Matches  168;  Conservative    0;  Mismatches   95;  Indels    6;  Gaps    2;

Qy        144 ATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTA 203
Db     125354 AATTTGAAGATGTCCTATACGGTCGATTGTTTCCAAAGACGACTGCAACACTCCATTCTA 125295

Qy        204 TCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGC 263
Db     125294 TC-----TTACATGCCCTTTTGCAATATGACTTTTCCACTTTTCCATCAAGACATAGACT 125240

Qy        264 TTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA CAATGGAATGT 322
Db     125239 CTATCTTACCTCCCCTTAAATCTGGACTAATCTTATGAGTTGCTTTAACCAATATCATGT 125180

Qy        323 AGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCAC 382
Db     125179 GGTAGAAATGACAATGTGCAACTCCAAAGACAAAACTTCAAGAATCCCCACACCTTCCAT 125120

Qy        383 TCACACTCTCTTGGAAACCAGATGCAATG 411
Db     125119 TTTTTATGTCTTGGAATGCAGCTGCCATG 125091


RESULT 12
AC022982/c
LOCUS       AC022982               63409 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-545P23 map 18, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC022982
VERSION     AC022982.2  GI:9163727
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 63409)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-545P23
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 63409)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6921670.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6494
                Center clone name: 545_P_23
            --------------.
            * NOTE: This record contains 73 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *         1     758: contig of 758 bp in length
            *       759     858: gap of 100 bp
            *       859    1633: contig of 775 bp in length
            *      1634    1733: gap of 100 bp
            *      1734    2499: contig of 766 bp in length
            *      2500    2599: gap of 100 bp
            *      2600    3355: contig of 756 bp in length
            *      3356    3455: gap of 100 bp
            *      3456    4210: contig of 755 bp in length
            *      4211    4310: gap of 100 bp
            *      4311    5031: contig of 721 bp in length
            *      5032    5131: gap of 100 bp
            *      5132    5902: contig of 771 bp in length
            *      5903    6002: gap of 100 bp
            *      6003    6781: contig of 779 bp in length
            *      6782    6881: gap of 100 bp
            *      6882    7661: contig of 780 bp in length
            *      7662    7761: gap of 100 bp
            *      7762    8519: contig of 758 bp in length
            *      8520    8619: gap of 100 bp
            *      8620    9396: contig of 777 bp in length
            *      9397    9496: gap of 100 bp
            *      9497   10268: contig of 772 bp in length
            *     10269   10368: gap of 100 bp
            *     10369   11144: contig of 776 bp in length
            *     11145   11244: gap of 100 bp
            *     11245   12015: contig of 771 bp in length
            *     12016   12115: gap of 100 bp
            *     12116   12882: contig of 767 bp in length
            *     12883   12982: gap of 100 bp
            *     12983   13754: contig of 772 bp in length
            *     13755   13854: gap of 100 bp
            *     13855   14624: contig of 770 bp in length
            *     14625   14724: gap of 100 bp
            *     14725   15482: contig of 758 bp in length
            *     15483   15582: gap of 100 bp
            *     15583   16336: contig of 754 bp in length
            *     16337   16436: gap of 100 bp
```

```
* 16437 17198: contig of 762 bp in length
* 17199 17298: gap of 100 bp
* 17299 18039: contig of 741 bp in length
* 18040 18139: gap of 100 bp
* 18140 18888: contig of 749 bp in length
* 18889 18988: gap of 100 bp
* 18989 19759: contig of 771 bp in length
* 19760 19859: gap of 100 bp
* 19860 20623: contig of 764 bp in length
* 20624 20723: gap of 100 bp
* 20724 21468: contig of 745 bp in length
* 21469 21568: gap of 100 bp
* 21569 22318: contig of 750 bp in length
* 22319 22418: gap of 100 bp
* 22419 23194: contig of 776 bp in length
* 23195 23294: gap of 100 bp
* 23295 24058: contig of 764 bp in length
* 24059 24158: gap of 100 bp
* 24159 24932: contig of 774 bp in length
* 24933 25032: gap of 100 bp
* 25033 25794: contig of 762 bp in length
* 25795 25894: gap of 100 bp
* 25895 26658: contig of 764 bp in length
* 26659 26758: gap of 100 bp
* 26759 27529: contig of 771 bp in length
* 27530 27629: gap of 100 bp
* 27630 28374: contig of 745 bp in length
* 28375 28474: gap of 100 bp
* 28475 29243: contig of 769 bp in length
* 29244 29343: gap of 100 bp
* 29344 30120: contig of 777 bp in length
* 30121 30220: gap of 100 bp
* 30221 30990: contig of 770 bp in length
* 30991 31090: gap of 100 bp
* 31091 31857: contig of 767 bp in length
* 31858 31957: gap of 100 bp
* 31958 32721: contig of 764 bp in length
* 32722 32821: gap of 100 bp
* 32822 33588: contig of 767 bp in length
* 33589 33688: gap of 100 bp
* 33689 34446: contig of 758 bp in length
* 34447 34546: gap of 100 bp
* 34547 35316: contig of 770 bp in length
* 35317 35416: gap of 100 bp
* 35417 36182: contig of 766 bp in length
* 36183 36282: gap of 100 bp
* 36283 37054: contig of 772 bp in length
* 37055 37154: gap of 100 bp
* 37155 37923: contig of 769 bp in length
* 37924 38023: gap of 100 bp
* 38024 38787: contig of 764 bp in length
* 38788 38887: gap of 100 bp
* 38888 39671: contig of 784 bp in length
* 39672 39771: gap of 100 bp
* 39772 40600: contig of 829 bp in length
* 40601 40700: gap of 100 bp
* 40701 41474: contig of 774 bp in length
* 41475 41574: gap of 100 bp
* 41575 42356: contig of 782 bp in length
* 42357 42456: gap of 100 bp
* 42457 43203: contig of 747 bp in length
* 43204 43303: gap of 100 bp
* 43304 44078: contig of 775 bp in length
* 44079 44178: gap of 100 bp
* 44179 44941: contig of 763 bp in length
* 44942 45041: gap of 100 bp
* 45042 45813: contig of 772 bp in length
* 45814 45913: gap of 100 bp
* 45914 46705: contig of 792 bp in length
* 46706 46805: gap of 100 bp
* 46806 47581: contig of 776 bp in length
* 47582 47681: gap of 100 bp
* 47682 48492: contig of 811 bp in length
* 48493 48592: gap of 100 bp
* 48593 49368: contig of 776 bp in length
* 49369 49468: gap of 100 bp
* 49469 50229: contig of 761 bp in length
* 50230 50329: gap of 100 bp
* 50330 51097: contig of 768 bp in length
* 51098 51197: gap of 100 bp
* 51198 51964: contig of 767 bp in length
* 51965 52064: gap of 100 bp
* 52065 52828: contig of 764 bp in length
* 52829 52928: gap of 100 bp
* 52929 53697: contig of 769 bp in length
* 53698 53797: gap of 100 bp
* 53798 54564: contig of 767 bp in length
* 54565 54664: gap of 100 bp
* 54665 55440: contig of 776 bp in length
* 55441 55540: gap of 100 bp
* 55541 56377: contig of 837 bp in length
* 56378 56477: gap of 100 bp
* 56478 57251: contig of 774 bp in length
* 57252 57351: gap of 100 bp
* 57352 58173: contig of 822 bp in length
* 58174 58273: gap of 100 bp
* 58274 59026: contig of 753 bp in length
* 59027 59126: gap of 100 bp
* 59127 59892: contig of 766 bp in length
* 59893 59992: gap of 100 bp
* 59993 60758: contig of 766 bp in length
* 60759 60858: gap of 100 bp
* 60859 61627: contig of 769 bp in length
```

```
Query Match            16.8%;  Score 84.4;  DB 2;  Length 63409;
Best Local Similarity  62.6%;  Pred. No. 1.1e-10;
Matches  164;  Conservative    0;  Mismatches   96;  Indels    2;  Gaps    2;

Qy        138 AAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTC 197
Db      24550 AAGTCTATTATATTATTACTCAGTATGGAAGGTTGTTTACCAAGGTGGCCAGAACAATTA 24491

Qy        198 CTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATG 257
Db      24490 TTCCCATCCCTATATCCAGGCACCTTGGCTGTGTGACTTCACCACTCCTCTCAAGAG-GA 24432

Qy        258 TGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCTTCTGACTTGCTTTGA-CAATG 316
Db      24431 AAGAGTCTATTTTACTTCCCCTTGAATTAAAACTGTCCTTGTGATTTGCTTTGACCAATA 24372

Qy        317 GAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACC 376
Db      24371 GAAGGTGTCACAGATGACATAGGACAACTTCTGGGCTTAGGTCTCAAGAGGCATCTCAAC 24312

Qy        377 TTCCACTCACACTCTCTTGGAA 398
Db      24311 ATCTGTTTTCAGCCCAATGGAA 24290
```

```
RESULT 13
AC008634
LOCUS       AC008634                97811 bp    DNA     linear   PRI 23-OCT 2001
DEFINITION  Homo sapiens chromosome 5 clone CTB-174K14, complete sequence.
ACCESSION   AC008634
VERSION     AC008634.5  GI:16328250
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 97811)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 97811)
  AUTHORS   DOE Joint Genome Institute.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 97811)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Oct 23, 2001 this sequence version replaced gi:9256017.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1..97811
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-174K14"
BASE COUNT    30500 a  19379 c  18444 g  29488 t
ORIGIN

  Query Match             16.8%;  Score 84.4;  DB 9;  Length 97811;
  Best Local Similarity   62.6%;  Pred. No. 1.1e-10;
  Matches  164;  Conservative    0;  Mismatches   96;  Indels    2;  Gaps    2;

Qy        138 AAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTC 197

Db      87692 AAGTCTATTATATTATTACTCAGTATGGAAGGTTGTTTACCAAGGTGGCCAGAACAATTA 87751

Qy        198 CTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATG 257

Db      87752 TTCCCATCCCTATATCCAGGCACCTTGGCTGTGTGACTTCACCACTCCTCTCAAGAG-GA 87810

Qy        258 TGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATG 316

Db      87811 AAGAGTCTATTTTACTTCCCCTTGAATTAAAACTGTCCTTGTGATTTGCTTTGACCAATA 87870

Qy        317 GAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACC 376

Db      87871 GAAGGTGTCACAGATGACACAGGACAACTTCTGGGCCTAGGTCTCAAGAGGCATCTCAAC 87930

Qy        377 TTCCACTCACACTCTCTTGGAA 398

Db      87931 ATCTGTTTTCAGCCCAATGGAA 87952


RESULT 14
AC011384/c
LOCUS       AC011384              104475 bp    DNA     linear   PRI 07-DEC-2001
DEFINITION  Homo sapiens chromosome 5 clone CTB-143D11, complete sequence.
ACCESSION   AC011384
VERSION     AC011384.4  GI:17402759
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104475)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104475)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 104475)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 7, 2001 this sequence version replaced gi:13752665.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1..104475
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-143D11"
BASE COUNT    31967 a  19178 c  19721 g  33609 t
ORIGIN

  Query Match             16.8%;  Score 84.4;  DB 9;  Length 104475;
  Best Local Similarity   62.6%;  Pred. No. 1e-10;
  Matches  164;  Conservative    0;  Mismatches   96;  Indels    2;  Gaps    2;

Qy        138 AAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTC 197

Db      94402 AAGTCTATTATATTATTACTCAGTATGGAAGGTTGTTTACCAAGGTGGCCAGAACAATTA 94343

Qy        198 CTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATG 257

Db      94342 TTCCCATCCCTATATCCAGGCACCTTGGCTGTGTGACTTCACCACTCCTCTCAAGAG-GA 94284

Qy        258 TGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATG 316

Db      94283 AAGAGTCTATTTTACTTCCCCTTGAATTAAAACTGTCCTTGTGATTTGCTTTGACCAATA 94224

Qy        317 GAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACC 376

Db      94223 GAAGGTGTCACAGATGACACAGGACAACTTCTGGGCCTAGGTCTCAAGAGGCATCTCAAC 94164

Qy        377 TTCCACTCACACTCTCTTGGAA 398

Db      94163 ATCTGTTTTCAGCCCAATGGAA 94142


RESULT 15
AL137799
LOCUS       AL137799              121394 bp    DNA     linear   HTG 10-JUL-2001
DEFINITION  Homo sapiens chromosome 1 clone RP5-1199A8, *** SEQUENCING IN
            PROGRESS ***, 6 unordered pieces.
ACCESSION   AL137799
VERSION     AL137799.4  GI:10039491
KEYWORDS    HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Plumb,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Sep 8, 2000 this sequence version replaced gi:9211917.
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: dJ1199A8
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
```

```
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 119318 bases at least Q40
            Consensus quality: 119968 bases at least Q30
            Consensus quality: 120408 bases at least Q20
            Insert size: 120894; sum-of-contigs
            Insert size: 128726; 9.8% error; agarose-fp
            Quality coverage: 6.06x in Q20 bases; sum-of-contigs Quality
            coverage: 5.79x in Q20 bases; agarose-fp
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1     54341: contig of 54341 bp in length
            *      54342     54441: gap of 100 bp
            *      54442     60268: contig of 5827 bp in length
            *      60269     60368: gap of 100 bp
            *      60369     90782: contig of 30414 bp in length
            *      90783     90882: gap of 100 bp
            *      90883    102742: contig of 11860 bp in length
            *     102743    102842: gap of 100 bp
            *     102843    107284: contig of 4442 bp in length
            *     107285    107384: gap of 100 bp
            *     107385    121394: contig of 14010 bp in length.
FEATURES             Location/Qualifiers
     source          1..121394
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP5-1199A8"
                     /clone_lib="RPCI-5"
     misc_feature    1..54341
                     /note="assembly_fragment:00965
                     fragment_chain:1
                     clone_end:SP6
                     vector_side:left"
     misc_feature    54442..60268
                     /note="assembly_fragment:01389
                     fragment_chain:1"
     misc_feature    60369..90782
                     /note="assembly_fragment:01706
                     fragment_chain:1"
     misc_feature    90883..102742
                     /note="assembly_fragment:00398
                     fragment_chain:2"
     misc_feature    102843..107284
                     /note="assembly_fragment:00472
                     fragment_chain:2"
     misc_feature    107385..121394
                     /note="assembly_fragment:01054
                     fragment_chain:2"
BASE COUNT    31229 a  31246 c  29603 g  28816 t    500 others
ORIGIN

  Query Match            16.8%;  Score 84.2;  DB 2;  Length 121394;
  Best Local Similarity   65.7%;  Pred. No. 1.1e-10;
  Matches  153;  Conservative    0;  Mismatches   78;  Indels    2;  Gaps    2;

Qy        216 TGCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCA 274

Db     119265 TGTCCCAGTACAATGTGACTTTGCTGCTTCTGTCATTAAGAGATGGAGTCTGTTTCCCCA 119324

Qy        275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 333

Db     119325 CCTCTTGAATCTGGGCTGATCTTATGATTTGTTTTGACCAGTAAAGTCTGGTGGAAGGGA 119384

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCT 393

Db     119385 TGCTGTGCAACTTCTGCATGTAGGCCTCAAGAAGCCTCGCAACTTCCACGCTCATTTTGG 119444

Qy        394 TGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATG 446

Db     119445 AGCTCTTCTGCGGCTCTGTGAAGAAGCCTGGACCAACCTGCTGGAGACATGTG 119497

Search completed: October 24, 2003, 21:06:02
Job time : 1389.29 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 126.251 Seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaaagtgctgggattacag.....tgaatgaggctagctaggcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	501	100.0	110000	22	AAF84800	Nucleotide sequenc
c	2	82.8	16.5	503	20	AAZ22641	Novel cDNA sequenc
c	3	82.6	16.5	854	20	AAZ22642	Novel nucleotide s
	4	82.6	16.5	27869	22	ABA19635	Human nervous syst
c	5	82.6	16.5	27869	22	AAK66517	Human immune/haema
	6	70	14.0	218336	25	ABQ76678	Androgen receptor
c	7	64.6	12.9	5239	24	ABL32273	Human immune syste
c	8	63	12.6	1241	21	AAC74337	Human secreted pro
	9	61.4	12.3	1479	22	AAD05322	Human secreted pro
	10	61.4	12.3	1484	22	AAD05372	Human secreted pro
	11	61	12.2	670	22	AAS46911	Human G protein-co
	12	61	12.2	670	24	ABK81689	cDNA encoding nove
	13	60.4	12.1	5239	24	ABL32272	Human immune syste
c	14	60.2	12.0	651	20	AAV88447	EST clone GI51. H
c	15	57.4	11.5	2477	22	AAH17565	Human cDNA sequenc
	16	56.8	11.3	1040	21	AAA81696	N. meningitidis pa
	17	56.6	11.3	567571	25	AAD53224	Human chromosome 3
	18	56.4	11.3	1757	25	ABX12157	cDNA encoding huma
c	19	56.4	11.3	5434	22	AAK76210	Human immune/haema
	20	53.8	10.7	3535	22	AAK65228	Human immune/haema
	21	53.8	10.7	3535	22	AAK73493	Human immune/haema
	22	53.6	10.7	1214	22	AAI59283	Human polynucleoti
	23	53.6	10.7	1275	22	AAI61069	Human polynucleoti
	24	53.6	10.7	1586	21	AAZ86793	Human protein kina
	25	53.6	10.7	1586	22	AAD11844	Human protein kina
	26	53.6	10.7	1586	24	ABQ76287	Human protein kina
	27	53.6	10.7	1868	24	AAS17048	Human protein kina
	28	53.6	10.7	1868	25	ABX14972	Human protein kina
	29	53.6	10.7	2142	22	AAH42064	Cell cycle protein
c	30	53.6	10.7	2275	22	AAS01478	Human secreted pro
	31	53.6	10.7	2538	22	AAH76219	Human kinase PKIN-
	32	53.6	10.7	3240	22	AAF75345	Human TGF-beta rec
	33	53.6	10.7	3387	22	AAF44687	Novel protein kina
	34	53.2	10.6	90220	24	ABX83576	Human cDNA differe
c	35	53.2	10.6	100000	24	ABQ74541	Human transglutami
c	36	52.6	10.5	397658	25	ABQ83210	Human transporter
	37	52.2	10.4	2141	22	AAH42063	Cell cycle protein
	38	51.6	10.3	245	21	AAC31581	Human secreted pro
	39	51.6	10.3	1434	22	AAK89069	Human digestive sy
	40	51.6	10.3	1434	22	AAI57626	Human colorectal c
	41	51.6	10.3	1434	24	ABS99803	Genomic DNA #7 enc
	42	51	10.2	502	23	ABV05821	Human prostate exp
c	43	50.4	10.1	10137	22	ABA18984	Human nervous syst
	44	50.2	10.0	389	22	AAK64465	Human immune/haema
c	45	50.2	10.0	972	22	AAK73675	Human immune/haema

ALIGNMENTS

RESULT 1
AAF84800
ID AAF84800 standard; DNA; 110000 BP.
XX
AC AAF84800;
XX
DT 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 gene.
XX
KW Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW autosomal dominant familial spastic paraplegia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 9932..102009
FT /*tag= a
FT /note= "contains introns"
FT exon 9932..10471
FT /*tag= b
FT /number= "1"
FT intron 10472..33718
FT /*tag= c
FT /number= "1"
FT exon 33719..33805
FT /*tag= d
FT /number= "2"
FT intron 33806..35748
FT /*tag= e
FT /number= "2"

```
FT   exon            35749..35832
FT                   /*tag= f
FT                   /number= "3"
FT   intron          35833..45022
FT                   /*tag= g
FT                   /number= "3"
FT   exon            45023..45118
FT                   /*tag= h
FT                   /number= "4"
FT   intron          45119..60863
FT                   /*tag= i
FT                   /number= "4"
FT   exon            60864..61051
FT                   /*tag= j
FT                   /number= "5"
FT   intron          61052..61927
FT                   /*tag= k
FT                   /number= "5"
FT   exon            61928..62061
FT                   /*tag= l
FT                   /number= "6"
FT   intron          62062..62344
FT                   /*tag= m
FT                   /number= "6"
FT   exon            62345..62438
FT                   /*tag= n
FT                   /number= "7"
FT   intron          62439..73173
FT                   /*tag= o
FT                   /number= "7"
FT   exon            73174..73248
FT                   /*tag= p
FT                   /number= "8"
FT   intron          73249..74633
FT                   /*tag= q
FT                   /number= "8"
FT   exon            74634..74705
FT                   /*tag= r
FT                   /number= "9"
FT   intron          74706..82788
FT                   /*tag= s
FT                   /number= "9"
FT   exon            82789..82864
FT                   /*tag= t
FT                   /number= "10"
FT   intron          82865..83102
FT                   /*tag= u
FT                   /number= "10"
FT   exon            83103..83194
FT                   /*tag= v
FT                   /number= "11"
FT   intron          83195..83334
FT                   /*tag= w
FT                   /number= "11"
FT   exon            83335..83414
FT                   /*tag= x
FT                   /number= "12"
FT   intron          83415..88129
FT                   /*tag= y
FT                   /number= "12"
FT   exon            88130..88172
FT                   /*tag= z
FT                   /number= "13"
FT   intron          88173..89561
FT                   /*tag= aa
FT                   /number= "13"
FT   exon            89562..89641
FT                   /*tag= ab
FT                   /number= "14"
FT   intron          89642..91162
FT                   /*tag= ac
FT                   /number= "14"
FT   exon            91163..91233
FT                   /*tag= ad
FT                   /number= "15"
FT   intron          91234..93443
FT                   /*tag= ae
FT                   /number= "15"
FT   exon            93444..93484
FT                   /*tag= af
FT                   /number= "16"
FT   intron          93485..100599
FT                   /*tag= ag
FT                   /number= "16"
FT   exon            100600..102009
FT                   /*tag= ah
FT                   /number= "17"
XX
PN   FR2798138-A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J, Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening -
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (FSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of FSF-AD.
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 other;

  Query Match          100.0%;  Score 501;  DB 22;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 8.1e-137;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4000 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 4059

Qy         61 CTTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4060 CTTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 4119

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4120 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAA 4179

Qy        181 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4180 GATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCT 4239

Qy        241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4240 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTG 4299

Qy        301 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       4300 ACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGTTT 4359

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 420

Db       4360 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGT 4419

Qy        421 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 480

Db       4420 CAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAACCTCTGAACAT 4479

Qy        481 ATGAATGAGGCTAGCTAGGCC 501

Db       4480 ATGAATGAGGCTAGCTAGGCC 4500


RESULT 2
AAZ22641/c
ID   AAZ22641 standard; cDNA; 502 BP.
XX
AC   AAZ22641;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Novel cDNA sequence 1 derived from the b2HFLS20W cDNA library.
XX
KW   interferon; novel sequence; pathogenic infection; antiviral; cell fate;
KW   cell proliferation; immune disorder; leukemia; wound healing; antibody;
KW   assay; ds.
XX
OS   Homo sapiens.
XX
PN   WO9945117-A2.
XX
PD   10-SEP-1999.
XX
PF   04-MAR-1999;   99WO-US03727.
XX
PR   04-MAR-1998;   98US-0034878.
PR   20-AUG-1998;   98US-0137348.
XX
PA   (HYSE-) HYSEQ INC.
XX
PT   Drmanac RT,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Ford J,  Mulero JJ;
XX
DR   WPI; 1999-551041/46.
XX
PT   New interferon polynucleotides useful for treating disorders involving
PT   pathogenic infection, cell fate and differentiation, inflammation etc.
PT   -
XX
PS   Claim 1; Fig 2; 94pp; English.
XX
CC   This a novel sequence isolated from the b2HFLS20W cDNA library, prepared
CC   from human fetal liver-spleen tissue, with another novel sequence
CC   (AAZ22642). The polypeptide sequences corresponding to these nucleic
CC   acid sequences have a striking homology to the interferon family of
CC   polypeptides.
CC   The interferon polypeptides may be administered therapeutically to
CC   treat/prevent disorders involving pathogenic infection (e.g. viral
CC   and bacterial infections) or cell fate and differentiation (e.g. immune
CC   disorders, leukemias, wound healing, tissue growth etc.). Interferons
CC   are known to have antiviral activity and are believed to be involved in
CC   regulatory processes such as cell proliferation and differentiation.
CC   The polypeptides may also inhibit platelet aggregation and therefore be
CC   used to treat/prevent conditions involving inflammatory responses. They
CC   can be combined with a carrier in compositions which can be administered
CC   to treat disorders as above, and used to generate antibodies useful in
CC   polypeptide detection and purification. They can also be used to
CC   identify binding compounds, by contacting with a compound and detecting
CC   complex formation, optionally within a cell in which the complex drives
CC   expression of a reporter gene. They (and the polynucleotides) are
CC   therefore useful in drug screening assays e.g. for the above disorders.
CC   The polynucleotides are also useful to detect polynucleotides encoding
CC   the polypeptides e.g. to diagnose disorders (or susceptibility to
CC   disorders) relating to polypeptide expression, or mark polypeptide-
CC   expressing tissues. They may also be used to produce probes useful to
CC   detect related sequences or for gene mapping.
XX
SQ   Sequence 502 BP; 141 A; 97 C; 143 G; 118 T; 3 other;

  Query Match             16.5%;  Score 82.8;  DB 20;  Length 502;
  Best Local Similarity   63.8%;  Pred. No. 2.1e-14;
  Matches  157;  Conservative    2;  Mismatches   87;  Indels    2;  Gaps    2;

Qy        203 ATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGA 261

Db        347 ATTCTCCCCATCCTGTCCCAGTACAATGTGACTTTGCTGCTTCTGTCATTAAGAGATGGA 288

Qy        262 GCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAAT 320

Db        287 GTCTGTTTCCCCACCTCTTGAATCTGGGCTGATCTTATGATTTGTTTTGACCAGTAAAGT 228

Qy        321 GTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCC 380

Db        227 CTGGTGGAAGGGATGCTGCGCAACTTCTGCATGTAGGCCTCAAGAAGCCTCGCAACTTCC 168

Qy        381 ACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGA 440

Db        167 ACGCTCATTTTGGAGCTCTTCTGCGGCTCTGTGAAGAAGCCTGGACCAACCTGCTGGAGA 108

Qy        441 CATATG 446

Db        107 CATGTG 102


RESULT 3
AAZ22642/c
ID   AAZ22642 standard; cDNA; 854 BP.
XX
AC   AAZ22642;
XX
DT   08-DEC-1999  (first entry)
XX
DE   Novel nucleotide sequence IFN-HY from the b2HFLS20W cDNA library.
XX
KW   interferon; novel sequence; pathogenic infection; antiviral; cell fate;
KW   cell proliferation; immune disorder; leukemia; wound healing; antibody;
KW   assay; ds.
XX
OS   Homo sapiens.
XX
PN   WO9945117-A2.
XX
PD   10-SEP-1999.
XX
PF   04-MAR-1999;   99WO-US03727.
XX
PR   04-MAR-1998;   98US-0034878.
PR   20-AUG-1998;   98US-0137348.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Ford J,  Mulero JJ;
XX
DR   WPI; 1999-551041/46.
DR   P-PSDB; AAY42442.
XX
PT   New interferon polynucleotides useful for treating disorders involving
PT   pathogenic infection, cell fate and differentiation, inflammation etc.
PT   -
XX
PS   Claim 1; Fig 2; 94pp; English.
XX
CC   This a novel sequence isolated from the b2HFLS20W cDNA library, prepared
```

```
CC   from human fetal liver-spleen tissue, with another novel sequence
CC   (AAZ22641). The polypeptide sequences corresponding to these nucleic
CC   acid sequences have a striking homology to the interferon family of
CC   polypeptides and this nucleotide sequence has been designated IFN-HY.
CC   The interferon polypeptides may be administered therapeutically to
CC   treat/prevent disorders involving pathogenic infection (e.g. viral
CC   and bacterial infections) or cell fate and differentiation (e.g. immune
CC   disorders, leukemias, wound healing, tissue growth etc.). Interferons
CC   are known to have antiviral activity and are believed to be involved in
CC   regulatory processes such as cell proliferation and differentiation.
CC   The polypeptides may also inhibit platelet aggregation and therefore be
CC   used to treat/prevent conditions involving inflammatory responses. They
CC   can be combined with a carrier in compositions which can be administered
CC   to treat disorders as above, and used to generate antibodies useful in
CC   polypeptide detection and purification. They can also be used to
CC   identify binding compounds, by contacting with a compound and detecting
CC   complex formation, optionally within a cell in which the complex drives
CC   expression of a reporter gene. They (and the polynucleotides) are
CC   therefore useful in drug screening assays e.g. for the above disorders.
CC   The polynucleotides are also useful to detect polynucleotides encoding
CC   the polypeptides e.g. to diagnose disorders (or susceptibility to
CC   disorders) relating to polypeptide expression, or mark polypeptide-
CC   expressing tissues. They may also be used to produce probes useful to
CC   detect related sequences or for gene mapping.
XX
SQ   Sequence 854 BP; 251 A; 180 C; 243 G; 180 T; 0 other;

  Query Match             16.5%;  Score 82.6;  DB 20;  Length 854;
  Best Local Similarity   65.2%;  Pred. No. 3e-14;
  Matches  152;  Conservative    0;  Mismatches   79;  Indels    2;  Gaps    2;

Qy        216 TGCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCA 274
                | |||   |||||| |||  ||  ||||   ||   ||    ||   |   |   ||
Db        334 TGTCCCAGTACAATGTGACTTTGCTGCTTCTGTCATTAAGAGATGGAGTCTGTTTCCCCA 275

Qy        275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 333
                 | |||    |  ||   || ||| |||  ||| | | | |  |   ||  ||
Db        274 CCTCTTGAATCTGGGCTGATCTTATGATTTGTTTTGACCAGTAAAGTCTGGTGGAAGGGA 215

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCT 393
                 || ||  ||    |  |  |||   ||   ||   |||   |  |
Db        214 TGCTGCGCAACTTCTGCATGTAGGCCTCAAGAAGCCTCGCAACTTCCACGCTCATTTTGG 155

Qy        394 TGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATG 446
                    |   ||   || | ||| | ||  | |   |||  |||  |
Db        154 AGCTCTTCTGCGGCTCTGTGAAGAAGCCTGGACCAACCTGCTGGAGACATGTG 102


RESULT 4
ABA19635
ID   ABA19635 standard; DNA; 27869 BP.
XX
AC   ABA19635;
XX
DT   23-JAN-2002  (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 11966.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01334.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
```

PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241819.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0252719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 11966; 1701pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABB14678-ABB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 27869 BP; 7054 A; 6673 C; 6322 G; 7820 T; 0 other;

Query Match 16.5%; Score 82.6; DB 22; Length 27869;
Best Local Similarity 65.2%; Pred. No. 1.1e-13;
Matches 152; Conservative 0; Mismatches 79; Indels 2; Gaps 2;

Qy 216 TGCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCA 274
Db 27536 TGTCCCAGTACAATGTGACTTTGCTGCTTCTGTCATTAAGAGATGGAGTCTGTTTCCCCA 27595

Qy 275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 332
Db 27596 CCTCTTGAATCTGGGCTGATCTTATGATTTGTTTTGACCAGTAAAGTCTGGTGGAAGGGA 27655

Qy 334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCT 393
Db 27656 TGCTGTGCAACTTCTGCATGTAGGCCTCAAGAAGCCTCGCAACTTCCACGCTCATTTTGG 27715

Qy 394 TGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATG 446
Db 27716 AGCTCTTCTGCGGCTCTGTGAAGAAGCCTGGACCAATCTGCTGGAGACATGTG 27768

RESULT 5
AAK66517/c
ID AAK66517 standard; DNA; 27869 BP.
XX
AC AAK66517;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21329.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-483426/52.
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -
XX
PS Disclosure; SEQ ID NO 21329; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I)

```
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 27869 BP; 7820 A; 6322 C; 6673 G; 7054 T; 0 other;

  Query Match             16.5%;  Score 82.6;  DB 22;  Length 27869;
  Best Local Similarity   65.2%;  Pred. No. 1.1e-13;
  Matches  152;  Conservative    0;  Mismatches   79;  Indels    2;  Gaps    2;

Qy        216 TGCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCA 274
Db        334 TGTCCCAGTACAATGTGACTTTGCTGCTTCTGTCATTAAGAGATGGAGTCTGTTTCCCCA 275

Qy        275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 333
Db        274 CCTCTTGAATCTGGGCTGATCTTATGATTTGTTTTGACCAGTAAAGTCTGGTGGAAGGGA 215

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCT 393
Db        214 TGCTGTGCAACTTCTGCATGTAGGCCTCAAGAAGCCTCGCAACTTCCACGCTCATTTTGG 155

Qy        394 TGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATG 446
Db        154 AGCTCTTCTGCGGCTCTGTGAAGAAGCCTGGACCAATCTGCTGGAGACATGTG 102


RESULT 6
ABQ76678
ID   ABQ76678 standard; DNA; 218336 BP.
XX
AC   ABQ76678;
XX
DT   26-MAR-2003  (first entry)
XX
DE   Androgen receptor signalling pathway-associated DNA AF067844.
XX
KW   Androgen receptor; transactivation; modulator; Smad3; Smad4; Akt; TGF-B;
KW   signal transduction pathway; transforming growth factor-B; phosphatase;
KW   tensin; cytostatic; antiproliferative; cellular proliferation; cancer;
KW   AF067844; ds.
XX
OS   Synthetic.
XX
PN   WO200282081-A2.
XX
PD   17-OCT-2002.
XX
PF   05-APR-2002; 2002WO-US11086.
XX
PR   06-APR-2001; 2001US-282266P.
PR   13-MAR-2002; 2002US-365060P.
XX
PA   (UYRP ) UNIV ROCHESTER.
XX
PI   Chang C;
XX
DR   WPI; 2003-046871/04.
XX
PT   Modulating androgen receptor activity, by administering a compound that
PT   modulates receptor activity, inhibits receptor-signal transduction
PT   pathway/receptor-coactivator interaction or changes amount or receptor
PT   -
XX
PS   Disclosure; Page 241-299; 302pp; English.
XX
CC   This invention describes a novel method for modulating androgen receptor
CC   activity or androgen receptor-mediated transactivation activity in a
CC   cell. The method involves administering a compound which causes
CC   modulation of the androgen receptors activity and the inhibition of
CC   interaction between the receptor and a protein involved in a signal
CC   transduction pathway. The compound also inhibits the interaction between
CC   the androgen receptor and a protein selected from Smad3, Smad4, Akt,
CC   transforming growth factor (TGF)-B and phosphatase and tensin homologues
CC   deleted on chromosome 10 (PTEN) or their fragments. The compounds of the
CC   invention have cytostatic and antiproliferative activity. The obtained
CC   composition is useful for treating any disease, where uncontrolled
CC   proliferation or cellular proliferation occurs such as cancer, e.g.
CC   prostate cancer. This sequence represents the androgen receptor
CC   transactivation signalling pathway modulator AF067844 described in
CC   the method of the invention.
XX
SQ   Sequence 218336 BP; 64194 A; 39437 C; 43295 G; 71406 T; 4 other;

  Query Match             14.0%;  Score 70;  DB 25;  Length 218336;
  Best Local Similarity   59.6%;  Pred. No. 1.2e-09;
  Matches  118;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy        259 GGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGA 318
Db      15917 GGAGTATCTTTCCCTTCCTTTTGAATCTGGGCTGTCCTTGTGACTTGCTTGGCCAAGAGA 15976

Qy        319 ATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTT 378
Db      15977 ATGCAGTGGAAGTGACATTCTGGGACTTCTGAACCCAGGATTTAAGAGACCCTGTATTTT 16036

Qy        379 CCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGA 438
Db      16037 CTGCTGTCTTCCTCTTGGGATCCAGGCACTGTATAAAGAAGCCTGGGTTGGACCACTGAA 16096

Qy        439 GACATATGTCCCAGCTAA 456
Db      16097 TAATGAGCAACCACACAA 16114


RESULT 7
ABL32273/c
ID   ABL32273 standard; DNA; 5239 BP.
XX
AC   ABL32273;
XX
DT   26-MAR 2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 246.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
```

```
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation
XX
PS   Claim 1; SEQ ID NO 246; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5239 BP; 1641 A; 84 C; 1235 G; 2379 T; 0 other;

  Query Match             12.9%;  Score 64.6;  DB 24;  Length 5239;
  Best Local Similarity   51.3%;  Pred. No. 1.2e-08;
  Matches  174;  Conservative    0;  Mismatches  164;  Indels    1;  Gaps    1;

Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db       3273 TTTAATAACTTTAAAAAACACCCAAAATTTAATAATTTAAATATTCAATCAAATTAAACT 3214

Qy        122 TCTGAACATGAGGTAGAAGGAAATTGTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db       3213 TCATTACGTTCTAATAACCAACAACCTAAAAATATTTATTACCATAAAAAAACTTACAAA 3154

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db       3153 ATATAACCATCAATCTCTCCTATTCTAATATACGTACCCCTTTACCGTATAACTTTACCA 3094

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db       3093 TTCCTCCTATCAAAAAATAAATTCTATACGTCCAATCTTAAATCTAAACTAACGTTATAA 3034

Qy        302 CTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       3033 TTTACTTTAACCAATAAAATATAACAAAAATAATATTATATAACTTTTAAAACTAAACCT 2974

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAA 399
Db       2973 CGAAAAACCTTACAACTTATATTTTAAATTCCTCAAAAA 2935


RESULT 8
AAC74337/c
ID   AAC74337 standard; cDNA; 1241 BP.
XX
AC   AAC74337;
XX
DT   02-FEB-2001  (first entry)
XX
DE   Human secreted protein gene 1 SEQ ID NO:11.
XX
KW   Human; secreted protein; diagnosis; cytostatic; immunosuppressive;
KW   nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW   antidiabetic; antiinflammatory; antiulcer; vulnerary; anticonvulsant;
KW   antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW   food additive; preservative; chromosome identification; cancer;
KW   female reproductive system disorder; immune disorder; wound healing;
KW   cardiovascular disorder; neurological disease; infectious disease;
KW   infection; ss.
XX
OS   Homo sapiens.
XX

PN   WO200058340-A2.
XX
PD   05-OCT-2000.
XX
PF   23-MAR-2000; 2000WO-US07724.
XX
PR   26-MAR-1999;   99US-0126510.
PR   07-JAN-2000; 2000US-0174850.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis G;
XX
DR   WPI; 2000-594638/56.
DR   P-PSDB; AAB39402.
XX
PT   Fifty nucleic acid molecules encoding human secreted proteins, useful
PT   in the prevention, treatment and diagnosis of cancer, immune disorders,
PT   cardiovascular disorders and neurological diseases
XX
PS   Claim 1; Page 329; 391pp; English.
XX
CC   The polynucleotide sequences given in AAC74337 to AAC74386 encode the
CC   human secreted proteins given in AAB39402 to AAB39451. AAB39452 to
CC   AAB39484 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Example of activities include:
CC   cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
CC   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC   cardiant. The polynucleotides and polypeptides are useful for preventing,
CC   treating or ameliorating a medical condition in e.g. humans, mice,
CC   rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC   can also be used as a food additive or preservative to increase or
CC   decrease storage capabilities. The polynucleotide are useful for
CC   chromosome identification. They are also useful as probes for diagnosing
CC   a disorder related to the female reproductive system, particularly breast
CC   and/or ovary cancer. They are also useful in the gene therapy of breast
CC   and ovarian cancer. Secreted protein nucleic acids, proteins,
CC   antibodies, agonists and antagonists are useful in the diagnosis,
CC   treatment and prevention of: (a) cancer; (b) immune disorders; (c)
CC   cardiovascular disorders; (d) wound healing; (e) neurological diseases;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections. AAC74328 to AAC74336 and AAB39401 represent
CC   sequences used in the exemplification of the present invention.
XX
SQ   Sequence 1241 BP; 363 A; 266 C; 257 G; 355 T; 0 other;

  Query Match             12.6%;  Score 63;  DB 21;  Length 1241;
  Best Local Similarity   67.4%;  Pred. No. 2e-08;
  Matches  118;  Conservative    0;  Mismatches   55;  Indels    2;  Gaps    2;

Qy        146 TCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAA-CAATTCCTCCTAT 204
Db        803 TATTCACATTCTTAATGTGGAAGTTTATTAGCAAAGATGGCTACAATTAATTTCTTCCAT 744

Qy        205 CCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTT-CTCTATCAAGATGTGGAGC 263
Db        743 TTCCATACATATACCATTTTATAATATGGCTCTGCCACTTGCCCCATCAAGAGATGGAAT 684

Qy        264 TTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGA 318
Db        683 TTATTTCTCCATACCATGAATCAGGGGTGGTCATGTAACATGCTTTGGCCAACGA 629


RESULT 9
AAD05322
ID   AAD05322 standard; cDNA; 1479 BP.
XX
AC   AAD05322;
XX
DT   17-JUL-2001  (first entry)
```

```
XX
DE   Human secreted protein-encoding gene 23 cDNA clone HISCV60, SEQ ID NO:33.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             19..405
FT                   /*tag= a
FT                   /product= "Human secreted protein"
FT   sig_peptide     19..96
FT                   /*tag= b
FT   mat_peptide     97..402
FT                   /*tag= c
FT                   /product= "Human mature secreted protein"
XX
PN   WO200134626-A1.
XX
PD   17-MAY-2001.
XX
PF   01-NOV-2000; 2000WO-US30045.
XX
PR   05-NOV-1999;   99US-0163581.
PR   30-JUN-2000; 2000US-0215133.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Moore PA,  Birse CE,  Ni J;
XX
DR   WPI; 2001-308778/32.
DR   P-PSDB; AAE01458.
XX
PT   New nucleic acid molecules encoding 28 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 439-440; 562pp; English.
XX
CC   AAD05330-AAD05379 represent cDNAs corresponding to 28 human secreted
CC   protein genes, and AAE01436-AAE01513 represent the proteins they encode.
CC   AAE01514-AAE01544 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   28 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 1479 BP; 420 A; 340 C; 297 G; 422 T; 0 other;

  Query Match             12.3%;  Score 61.4;  DB 22;  Length 1479;
  Best Local Similarity   57.2%;  Pred. No. 6.5e-08;
  Matches  131;  Conservative    0;  Mismatches   96;  Indels    2;  Gaps    1;

Qy        229 TGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGA 288

Db        642 TATGTGTTTGCCACCCTCCCATCAAGACATGGTGCCTATTTTTCATCCTTTGTGTCTGG 701

Qy        289 GTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTG 348

Db        702 GCTGGCCTTGTGTTCCATTTGACCAATAAGGTATGGTAGAAATGGCATTCTGGGACTTCT 761

Qy        349 GATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCA 408

Db        762 GAGCCCAACCTT--AAAAGACATTGAACTTTTGCTTTAGTGAGAGGGGAATCAAGATACC 819

Qy        409 ATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAAT 457

Db        820 CTGTAAAGAGGTCTGTGCTGTCCTTCTGAAGACAGAGGTCCTGGAAGAT 868


RESULT 10
AAD05372
ID   AAD05372 standard; cDNA; 1484 BP.
XX
AC   AAD05372;
XX
DT   17-JUL-2001  (first entry)
XX
DE   Human secreted protein-encoding gene 23 cDNA clone HISCV60, SEQ ID NO:83.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   inflammation; allergy; neurological disorder; Alzheimer's disease;
KW   Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW   skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW   cardiovascular disorder; angiogenic disorder; kidney disorder;
KW   gastrointestinal disorder; pregnancy-related disorder;
KW   endocrine disorder; infection; wound healing; vulnerary;
KW   cell culture; chemotaxis; food additive; gene therapy;
KW   binding partner identification; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             9..395
FT                   /*tag= a
FT                   /product= "Human secreted protein"
FT                   /transl_except= (pos:207..209, aa:Xaa)
FT                   /transl_except= (pos:210..212, aa:Xaa)
FT                   /transl_except= (pos:216..218, aa:Xaa)
FT                   /transl_except= (pos:294..296, aa:Xaa)
FT                   /note= "Xaa equals any of the twenty naturally ocurring
FT                   L-amino acids"
FT   sig_peptide     9..86
FT                   /*tag= b
FT   mat_peptide     87..392
FT                   /*tag= c
FT                   /product= "Human mature secreted protein"
XX
PN   WO200134626 A1.
XX
PD   17-MAY-2001.
XX
```

```
PF   01-NOV-2000; 2000WO-US30045.
XX
PR   05-NOV-1999;   99US-0163581.
PR   30-JUN-2000; 2000US-0215133.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Moore PA,  Birse CE,  Ni J;
XX
DR   WPI; 2001-308778/32.
DR   P-PSDB; AAE01506.
XX
PT   New nucleic acid molecules encoding 28 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 477-478; 562pp; English.
XX
CC   AAD05300-AAD05379 represent cDNAs corresponding to 28 human secreted
CC   protein genes, and AAE01436-AAE01513 represent the proteins they encode.
CC   AAE01514-AAE01544 represent human secreted protein fragments or variants.
CC   The genes and their secreted proteins are useful for preventing,
CC   treating or ameliorating medical conditions, e.g., by protein or gene
CC   therapy. Pathological conditions can be diagnosed by determining the
CC   amount of the new protein in a sample or by determining the presence of
CC   mutations in the new genes. Specific uses are described for each of the
CC   28 genes, based on the tissues in which they are most highly expressed,
CC   and include developing products for the diagnosis or treatment of
CC   proliferative disorders, cancer, tumours, foetal and developmental
CC   abnormalities, haematopoietic disorders, diseases of the immune system,
CC   AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC   allergies, neurological disorders (e.g., Alzheimer's disease,
CC   Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC   skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC   cardiovascular disorders, angiogenic disorders, kidney disorders,
CC   gastrointestinal disorders, pregnancy-related disorders, endocrine
CC   disorders, and infections. The proteins can also be used to aid wound
CC   healing and epithelial cell proliferation, to prevent skin aging due to
CC   sunburn, to maintain organs before transplantation, for supporting cell
CC   culture of primary tissues, to regenerate tissues, to identify their
CC   cognate ligands or binding partners, and in chemotaxis, and can be used
CC   as a food additive or preservative to modify storage properties.
CC   Antibodies specific for a protein of the invention can be used in
CC   alleviating symptoms associated with the disorders mentioned above, and
CC   in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC   immunosorbent assay (ELISA). The present sequence represents a human
CC   secreted protein-encoding cDNA of the invention.
XX
SQ   Sequence 1484 BP; 418 A; 337 C; 300 G; 422 T; 7 other;

  Query Match            12.3%;  Score 61.4;  DB 22;  Length 1484;
  Best Local Similarity  57.2%;  Pred. No. 6.5e-08;
  Matches  131;  Conservative    0;  Mismatches   96;  Indels    2;  Gaps    1;

Qy        229 TGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTCCCATATATTGCACTAGA 288
Db        632 TATGTGTTTGCCACCCTCCCATCAAGACATGGTGCCTATTTTTTCATCCTTTGTGTCTGG 691

Qy        289 GTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTG 348
Db        692 GCTGGCCTTGTGTTCCATTTGACCAATAAGGTATGGTAGAAATGGCATTCTGGGACTTCT 751

Qy        349 GATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCA 408
Db        752 GAGCCCAACCTT--AAAAGACATTGAACTTTTGCTTTAGTGAGAGGGGAATCAAGATACC 809

Qy        409 ATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAAT 457
Db        810 CTGTAAAGAGGTCTGTGCTGTCCTTCTGAAGACAGAGGTCCTGGAAGAT 858

RESULT 11
AAS46911

ID   AAS46911 standard; cDNA; 670 BP.
XX
AC   AAS46911;
XX
DT   19-DEC-2001  (first entry)
XX
DE   Human G protein-coupled receptor (GPCR) cDNA #93.
XX
KW   Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia;
KW   neurological disorder; metabolic disorder; cancer; rheumatoid arthritis;
KW   thyroid disorder; neurodegenerative disorder; cardiovascular disorder;
KW   renal failure; autoimmune disorder; hyperproliferative disorder; HIV; ss;
KW   human immunodeficiency virus; viral infection; neuroprotective;
KW   immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant;
KW   anorectic; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200168858-A2.
XX
PD   20-SEP-2001.
XX
PF   16-MAR-2001; 2001WO-US08456.
XX
PR   16-MAR-2000; 2000US-187783P.
PR   16-MAR-2000; 2000US-189907P.
PR   16-MAR-2000; 2000US-189917P.
PR   16-MAR-2000; 2000US-189916P.
PR   16-MAR-2000; 2000US-189960P.
PR   29-MAR-2000; 2000US-192155P.
PR   29-MAR-2000; 2000US-192234P.
PR   29-MAR-2000; 2000US-192830P.
PR   29-MAR-2000; 2000US-192916P.
PR   29-MAR-2000; 2000US-192923P.
PR   29-MAR-2000; 2000US-192933P.
PR   29-MAR-2000; 2000US-192945P.
XX
PA   (PHAA ) PHARMACIA & UPJOHN CO.
XX
PI   Vogeli G;
XX
DR   WPI; 2001-607458/69.
DR   P-PSDB; AAU29472.
XX
PT   Nucleic acid encoding G protein-coupled receptors, useful for the
PT   prevention, diagnosis and treatment of mental disorders -
XX
PS   Claim 4; Page 90; 274pp; English.
XX
CC   Sequences AAS46819-AAS46946 represent cDNA molecules encoding human G
CC   protein-coupled receptor (GPCR) polypeptides. The protein and DNA
CC   sequences of the invention can be used to identify compounds which bind
CC   to GPCR polypeptides and in screening for compounds that modulate GPCR
CC   activity. By screening a human subject for the presence of mutations in
CC   GPCR DNA, a GPCR-related disorder or a genetic predisposition can be
CC   diagnosed. The sequences can also be used for treatment and prevention of
CC   mental disorders such as schizophrenia, neurological disorders such as
CC   manic depression, metabolic disorders such as obesity, cancer, rheumatoid
CC   arthritis, thyroid disorders such as myxoedema, neurodegenerative
CC   disorders such as Parkinson's disease, cardiovascular disorders such as
CC   atherosclerosis, renal failure, autoimmune disorders, hyperproliferative
CC   disorders such as psoriasis and viral infections such as those caused by
CC   HIV.
XX
SQ   Sequence 670 BP; 218 A; 134 C; 143 G; 175 T; 0 other;

  Query Match            12.2%;  Score 61;  DB 22;  Length 670;
  Best Local Similarity  60.6%;  Pred. No. 6.4e-08;
  Matches  117;  Conservative    0;  Mismatches   75;  Indels    1;  Gaps    1;

Qy        237 TGCTACTTCTCTATCAAGATGTGGAGCTTATTTCCCATATATTGCACTAGAGTTGGCCT 296
Db        319 TGCAGTGTTTTCATCAAGAGCTGCAGTCTATTTCCCCACCCAAAATCCA-AGGTGGCCT 377
```

```
Qy        297 TCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAG 356
Db        378 TGTGACTTGCTTTGAAAATGAAATGCAGTGGAAGGAATTTTGTGTGACTTTCCAGACTAG 437

Qy        357 GTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAG 416
Db        438 GCTGCAAGAGACTTTGCAGATTCCCCTTCCACATTCTTGGAATGCTGCCCTGAGATAGCC 497

Qy        417 AAGTCAGGGCTAT 429
Db        496 ATGCAAGGAAGAT 510
```

RESULT 12
ABK81689
ID   ABK81689 standard; DNA; 670 BP.
XX
AC   ABK81689;
XX
DT   13-AUG-2002 (first entry)
XX
DE   cDNA encoding novel G protein coupled receptor (nGCPR-x) #93.
XX
KW   G protein coupled receptor; nGPCR-x; immune response; thyroid disorder;
KW   mental disorder; thyrectoxicosis; myxoedema; inflammatory condition;
KW   Crohn's disease; cell differentiation; homeostasis; rheumatoid arthritis;
KW   renal failure; autoimmune disorder; movement disorder; CNS disorder;
KW   viral infection; human immunodeficiency virus; HIV; metabolic disorder;
KW   cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy;
KW   porliferative disease; cancer; psoriasis; lung cancer; hormonal disorder;
KW   sexual dysfunction; gene; ss.
XX
OS   Homo sapiens.
XX
PN   US2002058306-A1.
XX
PD   16-MAY-2002.
XX
PF   16-MAR-2001; 2001US-0811284.
XX
PR   16-MAR-2000; 2000US-189783P.
PR   16-MAR-2000; 2000US-189907P.
PR   16-MAR-2000; 2000US-189917P.
PR   16-MAR-2000; 2000US-189918P.
PR   16-MAR-2000; 2000US-189960P.
PR   24-MAR-2000; 2000US-192155P.
PR   27-MAR-2000; 2000US-192234P.
PR   29-MAR-2000; 2000US-192830P.
PR   29-MAR-2000; 2000US-192945P.
PR   29-MAR-2000; 2000US-192916P.
PR   29-MAR-2000; 2000US-192923P.
PR   29-MAR-2000; 2000US-192830P.
PR   29-MAR-2000; 2000US-192945P.
PR   29-MAR-2000; 2000US-192830P.
PR   29-MAR-2000; 2000US-192945P.
PR   29-MAR-2000; 2000US-192830P.
PR   29-MAR-2000; 2000US-192945P.
XX
PA   (VOGE/) VOGELI G.
XX
PI   Vogeli G;
XX
DR   WPI; 2002-434856/46.
DR   P-PSDB; ABG60760.
XX
PT   New isolated nucleic acid encoding a G protein coupled receptor for
PT   producing the receptor which can induce an immune response in a mammal
PT   -
XX
PS   Claim 4; Page 65-66; 216pp; English.
XX
CC   The invention describes an isolated nucleic acid (I) comprising a
CC   sequence encoding a portion of a G protein coupled receptor (nGPCR-x).
CC   (I) is used to produce a recombinant nGPRC-x polypeptide. A polypeptide
CC   encoded by (I) is used to induce an immune response in a mammal. nGPRC-x
CC   is used to identify a compound that binds to it and/or modulates it's
CC   activity. (I) is used to identify animal homologues of nGPCR-x. (I) can
CC   be used to diagnose a human subject as having a brain or genetic
CC   predisposition disorder, such as a mental disorder. (I) is used to
CC   screen for an nGPCR-x related disorder including thyroid disorders (e.g.
CC   thyrectoxicosis, myxoedema), renal failure, inflammatory conditions (e.g.
CC   Crohn's disease), diseases related to cell differentiation and
CC   homeostasis, rheumatoid arthritis, autoimmune disorders, movement
CC   disorders, CNS disorders, viral infections (e.g. Human immunodeficiency
CC   virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity,
CC   anorexia, cardiomyopathies), porliferative diseases and cancers (e.g.
CC   psoriasis, lung cancer), hormonal disorders, sexual dysfunction and
CC   hereditary mental disorders in a human patient. A host cell comprising
CC   (I) is used to screen for a modulator of nGPCR-x activity. nGPCR-x is
CC   used to identify compounds that can treat mental disorders. The
CC   polypeptide encoded by (I) is used to purify a G protein from a sample.
CC   This sequence encodes a novel G protein coupled receptor (nGPCR-x)
CC   protein described in the invention.
XX
SQ   Sequence 670 BP; 218 A; 134 C; 143 G; 175 T; 0 other;

  Query Match   12.2%; Score 61; DB 24; Length 670;
  Best Local Similarity   60.6%; Pred. No. 6.4e-08;
  Matches 117; Conservative   0; Mismatches   75; Indels   1; Gaps   1;

```
Qy        237 TGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCT 296
Db        319 TGCAGTGTTTTCATCAAGAGCTGCAGTCTATTTCCCCACCCCAAAATCCA-AGGTGGCCT 377

Qy        297 TCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAG 356
Db        378 TGTGACTTGCTTTGAAAATGAAATGCAGTGGAAGGAATTTTGTGTGACTTTCCAGACTAG 437

Qy        357 GTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAG 416
Db        438 GCTGCAAGAGACTTTGCAGATTCCCCTTCCACATTCTTGGAATGCTGCCCTGAGATAGCC 497

Qy        417 AAGTCAGGGCTAT 429
Db        498 ATGCAAGGAAGAT 510
```

RESULT 13
ABL32272
ID   ABL32272 standard; DNA; 5239 BP.
XX
AC   ABL32272;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 245.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.

```
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation -
XX
PS   Claim 1; SEQ ID NO 245; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 5239 BP; 1265 A; 84 C; 1198 G; 2692 T; 0 other;

  Query Match            12.1%;  Score 60.4;  DB 24;  Length 5239;
  Best Local Similarity   50.6%;  Pred. No. 2e-07;
  Matches  171;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db       1967 TTTGATGGTTTTAAAAGGTATTTAAAGTTTAGTGATTTGAATGTTTAGTTAGGTTAGGTT 2026

Qy        122 TCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db       2027 TTATTACGTTTTGGTAATTAATAATTTAGAAATATTTGTTGTTATAGGAGGGTTTATAAA 2086

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db       2087 ATATAGTTATTAATTTTTTTTATTTTGATGTGTTTGTTTTTTTGTCGTGTGATTTTGTTA 2146

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db       2147 TTTTTTTTATTAAGAGGTAAATTTTATGTTTTTAGTTTAAAATTTGGGTTGATTTTGTGA 2206

Qy        302 CTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       2207 TTTGTTTTGATTAATAGAATGTGGTAGAAGTGATGTTATGTGATTTTTGGGGTTAGGTTT 2266

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAA 398
Db       2267 CGAGAGATTTTGTAGTTTATGTTTTGGATTTTTTAGAA 2304


RESULT 14
AAV88447/c
ID   AAV88447 standard; cDNA; 651 BP.
XX
AC   AAV88447;
XX
DT   12-FEB-1999  (first entry)
XX
DE   EST clone GI51.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845437-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;  98WO-US06956.
XX
PR   10-APR-1997;  97US-0837312.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI   Racie LA, Spaulding V, Treacy M;
XX
DR   WPI; 1999-070078/06.
XX
PT   New polynucleotides encoding human secreted proteins derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 400; 641pp; English.
XX
CC   The present sequence represents an expressed sequence tag (EST), and is
CC   a polynucleotide of the invention. The polynucleotides of the invention
CC   are all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 651 BP; 196 A; 127 C; 142 G; 186 T; 0 other;

  Query Match            12.0%;  Score 60.2;  DB 20;  Length 651;
  Best Local Similarity   61.6%;  Pred. No. 1.1e-07;
  Matches  146;  Conservative    0;  Mismatches   88;  Indels    3;  Gaps    3;

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db        368 AAGGCCACAAATGAGTAACTCCTTCCCATGTCCAGACTCCTTTGCAATGTGACCTTGCTA 309

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db        308 CTTCTC-ACCAGGAAATTGAGTCTATTTTTCCATTCTTTTGAATCTGGCTGGCCTTGTGA 250

Qy        302 CTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTT-TGGATTTTAGGTT 359
Db        249 CTTGCTTTGACTAACACAAGGTGGCAGAAGTATCACTGTATGAGTTCTTGGGCATAGGGC 190

Qy        360 TCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAG 416
Db        189 TTAAGAGGTATTGCAAGTTCTACTCTTTTCTTTTTGGGACCCTAAGCCATCATGATG 133


RESULT 15
AAH17555/c
ID   AAH17555 standard; cDNA; 2477 BP.
XX
AC   AAH17555;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human cDNA sequence SEQ ID NO:17036.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
```

PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0214036.
PR 27-AUG-1999; 99JP-0239253.
PR 11-JAN-2000; 2000JP-0003776.
PR 02-MAY-2000; 2000JP-0133767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17036; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonuclestide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonuclectide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH16742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2477 BP; 697 A; 416 C; 560 G; 804 T; 0 other;

Query Match 11.5%; Score 57.4; DB 22; Length 2477;
Best Local Similarity 59.5%; Pred. No. 1.2e-06;
Matches 97; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
Qy      210 TATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTT 269
Db      944 TATCTACAGCCTTTTGCAATGTGACTTTGCAGCTTTTCCATCAAGAGATCAAGTCTATTT 885

Qy      270 TCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGAGAATGGAATGTAGTACAA 329
Db      884 CCCCAGCCTATGAATCTGGACTTGCCTTGTGACTTTCTGTGAGCAACAGAATATAACGAA 825

Qy      330 ATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTA 372
Db      824 GTAACATTGTGCTGGTTCTGAGCCCAAACCTCAGGAGATCCTA 782
```

Search completed: October 24, 2003, 18:47:15
Job time : 128.251 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 12:34:49 ; Search time 31.3542 Seconds
(without alignments)
7052.739 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaaagtgctgggattacag..........tgaatgaggctagctaggcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	53.6	10.7	1586	3	US-09-173-581-11	Sequence 11, Appl
	2	53.6	10.7	1586	3	US-09-420-915-11	Sequence 11, Appl
	3	53.6	10.7	1868	4	US-09-387-212-1	Sequence 1, Appli
	4	53.6	10.7	1869	4	US-09-948-802-1	Sequence 1, Appli
	5	53.6	10.7	2142	4	US-09-441-039-3	Sequence 3, Appli
	6	52.2	10.4	2141	4	US-09-441-039-1	Sequence 1, Appli
	7	47.8	9.5	537	3	US-08-642-274D-38	Sequence 38, Appl
	8	47.8	9.5	537	3	US-08-952-014C-38	Sequence 38, Appl
c	9	46.6	9.3	465	4	US-09-499-522-5	Sequence 5, Appli
	10	45.6	9.1	63588	4	US-09-873-404-3	Sequence 3, Appli
c	11	45.2	9.0	162450	4	US-09-345-882-1	Sequence 1, Appli
c	12	45	9.0	3609	4	US-09-705-299-11	Sequence 11, Appl
	13	44.4	8.9	43950	4	US-09-735-934A-3	Sequence 3, Appli
	14	44.4	8.9	43950	4	US-10-060-332-3	Sequence 3, Appli
	15	44	8.8	17041	1	US-08-076-011-1	Sequence 1, Appli
	16	43.8	8.7	3350	3	US-09-110-116-2	Sequence 2, Appli
	17	43.8	8.7	17138	3	US-09-813-819-3	Sequence 3, Appli
	18	43.8	8.7	17138	4	US-09-920-048-3	Sequence 3, Appli
	19	43.8	8.7	64467	4	US-09-803-671B-3	Sequence 3, Appli
c	20	43.6	8.7	2373	3	US-08-975-762-45	Sequence 45, Appl
c	21	43.6	8.7	2373	3	US-09-295-028-45	Sequence 45, Appl
c	22	43.6	8.7	2373	4	US-09-106-582-45	Sequence 45, Appl
c	23	43.2	8.6	46718	4	US-09-816-093-3	Sequence 3, Appli
	24	43	8.6	41694	4	US-09-536-059-1	Sequence 1, Appli
	25	43	8.6	98844	4	US-09-791-211-10	Sequence 10, Appl
	26	42.8	8.5	99500	4	US-09-738-096-10	Sequence 10, Appl
c	27	42.6	8.5	14753	4	US-09-821-736-3	Sequence 3, Appli
c	28	42.2	8.4	35060	3	US-08-814-095-7	Sequence 7, Appli
	29	42.2	8.4	40000	4	US-09-780-049-18	Sequence 18, Appl
	30	42.2	8.4	65042	4	US-09-784-316-3	Sequence 3, Appli
c	31	42.2	8.4	84495	4	US-09-797-906-3	Sequence 3, Appli
	32	42	8.4	776	4	US-09-535-008-37	Sequence 37, Appl
	33	42	8.4	46718	4	US-09-816-093-3	Sequence 3, Appli
c	34	42	8.4	174493	4	US-09-804-471A-3	Sequence 3, Appli
	35	41.8	8.3	36651	4	US-09-738-894A-3	Sequence 3, Appli
	36	41.8	8.3	36651	4	US-09-964-469-3	Sequence 3, Appli
c	37	41.8	8.3	319608	4	US-09-539-333D-1	Sequence 1, Appli
c	38	41.8	8.3	319608	4	US-09-679-409-1	Sequence 1, Appli
	39	41.6	8.3	1811	1	US-08-849-252-1	Sequence 1, Appli
c	40	41.6	8.3	36651	4	US-09-738-894A-3	Sequence 3, Appli
c	41	41.6	8.3	36651	4	US-09-964-469-3	Sequence 3, Appli
	42	41.6	8.3	174493	4	US-09-804-471A-3	Sequence 3, Appli
	43	41.4	8.3	4823	2	US-08-457-254-5	Sequence 5, Appli
	44	41.4	8.3	4823	2	US-08-484-257-20	Sequence 20, Appl
	45	41.4	8.3	4823	3	US-08-999-927-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-173-581-11
; Sequence 11, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina
; APPLICANT: Azimzai, Yalda
; APPLICANT: Lu, Aina
; TITLE OF INVENTION: Protein Kinase Homologs
; FILE REFERENCE: PF-0614 US
; CURRENT APPLICATION NUMBER: US/09/173,581A
; CURRENT FILING DATE: 1998-10-15
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 1586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 132750
US-09-173-581-11

Query Match 10.7%; Score 53.6; DB 3; Length 1586;
Best Local Similarity 59.2%; Pred. No. 5.7e-07;
Matches 108; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

```
Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATC 251
              
Db       1154 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCCTCCCTTC 1213

Qy        252 AAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA 311
              
Db       1214 CATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTGA 1273

Qy        312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACT 370
              
Db       1274 TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCCT 1333

Qy        371 TACA 374
                |
Db       1334 TATA 1337
```

RESULT 2

```
US-09-420-915-11
; Sequence 11, Application US/09420915
; Patent No. 6264947
; GENERAL INFORMATION:
;  APPLICANT: Bandman, Olga
;  APPLICANT: Tang, Y. Tom
;  APPLICANT: Hillman, Jennifer L.
;  APPLICANT: Yue, Henry
;  APPLICANT: Guegler, Karl J.
;  APPLICANT: Corley, Neil C.
;  APPLICANT: Gorgone, Gina
;  APPLICANT: Azimzai, Yalda
;  APPLICANT: Lu, Aina
;  TITLE OF INVENTION: Protein Kinase Homologs
;  FILE REFERENCE: PF-0614 US
;  CURRENT APPLICATION NUMBER: US/09/420,915
;  CURRENT FILING DATE: 1999-10-20
;  EARLIER APPLICATION NUMBER: US 09/173,581
;  EARLIER FILING DATE: 1998-10-15
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PERL Program
; SEQ ID NO 11
;  LENGTH: 1586
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE: -
;  OTHER INFORMATION: 132750
US-09-420-915-11

  Query Match            10.7%;  Score 53.6;  DB 3;  Length 1586;
  Best Local Similarity  59.2%;  Pred. No. 5.7e-07;
  Matches  109;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATC 251
Db       1154 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCCTCCCTTC 1213

Qy        252 AAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA 311
Db       1214 CATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTGA 1273

Qy        312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACT 370
Db       1274 TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCCT 1333

Qy        371 TACA 374
Db       1334 TATA 1337


RESULT 3
US-09-387-212-1
; Sequence 1, Application US/09387212A
; Patent No. 6309849
; GENERAL INFORMATION:
;  APPLICANT:  ROBISON, KEITH E.
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
;  TITLE OF INVENTION:  PHOSPHATASE HOMOLOGUES AND USES THEREFOR
;  FILE REFERENCE:  MNI-090
;  CURRENT APPLICATION NUMBER: US/09/387,212A
;  CURRENT FILING DATE:  1999-08-31
;  NUMBER OF SEQ ID NOS:  18
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 1868
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: All occurences of n indicate any nucleotide
US-09-387-212-1

  Query Match            10.7%;  Score 53.6;  DB 4;  Length 1868;
  Best Local Similarity  59.2%;  Pred. No. 6.1e-07;
  Matches  109;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATC 251
Db       1415 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCCTCCCTTC 1474

Qy        252 AAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA 311
Db       1475 CATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTGA 1534

Qy        312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACT 370
Db       1535 TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCCT 1594

Qy        371 TACA 374
Db       1595 TATA 1598


RESULT 4
US-09-948-802-1
; Sequence 1, Application US/09948802
; Patent No. 6465232
; GENERAL INFORMATION:
;  APPLICANT: ROBISON, KEITH E.
;  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
;  TITLE OF INVENTION:  PHOSPHATASE HOMOLOGUES AND USES THEREFOR
;  FILE REFERENCE: MNI-090
;  CURRENT APPLICATION NUMBER: US/09/948,802
;  CURRENT FILING DATE:  2001-09-07
;  PRIOR APPLICATION NUMBER: 09/387,212
;  PRIOR FILING DATE: 1999-08-31
;  NUMBER OF SEQ ID NOS: 18
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;  LENGTH: 1868
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: All occurences of n indicate any nucleotide
US-09-948-802-1

  Query Match            10.7%;  Score 53.6;  DB 4;  Length 1868;
  Best Local Similarity  59.2%;  Pred. No. 6.1e-07;
  Matches  109;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATC 251
Db       1415 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCCTCCCTTC 1474

Qy        252 AAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA 311
Db       1475 CATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTGA 1534

Qy        312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACT 370
Db       1535 TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCCT 1594

Qy        371 TACA 374
Db       1595 TATA 1598


RESULT 5
US-09-441-039-3
; Sequence 3, Application US/09441039
; Patent No. 6428980
; GENERAL INFORMATION:
;  APPLICANT: Luo, Ying
;  APPLICANT: Yu, Pei Wen
;  APPLICANT: Shen, Mary
;  APPLICANT: Huang, Chao Bai
;  TITLE OF INVENTION: No. 6428980el RIP3 Associated Cell Cycle Proteins, Compositions
```

```
;  TITLE OF INVENTION: and Methods of Use
;  FILE REFERENCE: A68411/DJB/RMS/DAV
;  CURRENT APPLICATION NUMBER: US/09/441,039
;  CURRENT FILING DATE: 1999-11-16
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 2142
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-441-039-3

  Query Match            10.7%;  Score 53.6;  DB 4;  Length 2142;
  Best Local Similarity  59.2%;  Pred. No. 6.4e-07;
  Matches  109;  Conservative    0;  Mismatches   74;  Indels    1;  Gaps    1;

Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATC 251
Db       1698 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCCTCCCTTC 1757

Qy        252 AAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA 311
Db       1758 CATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTGA 1817

Qy        312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACT 370
Db       1818 TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCCT 1877

Qy        371 TACA 374
Db       1878 TATA 1881

RESULT 6
US-09-441-039-1
; Sequence 1, Application US/09441039
; Patent No. 6428980
; GENERAL INFORMATION:
;  APPLICANT: Luo, Ying
;  APPLICANT: Yu, Pei Wen
;  APPLICANT: Shen, Mary
;  APPLICANT: Huang, Chao Bai
;  TITLE OF INVENTION: No. 6428980el RIP3 Associated Cell Cycle Proteins, Compositions
;  TITLE OF INVENTION: and Methods of Use
;  FILE REFERENCE: A68411/DJB/RMS/DAV
;  CURRENT APPLICATION NUMBER: US/09/441,039
;  CURRENT FILING DATE: 1999-11-16
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 2141
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-441-039-1

  Query Match            10.4%;  Score 52.2;  DB 4;  Length 2141;
  Best Local Similarity  62.2%;  Pred. No. 1.7e-06;
  Matches  115;  Conservative    0;  Mismatches   68;  Indels    2;  Gaps    2;

Qy        192 CAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCT-CTAT 250
Db       1696 CAAATGTATCTTTCTGGGTATCCACACCTCTTGGCAATTGAAATTTGCAGCTCCTCCCTT 1755

Qy        251 CAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTG 310
Db       1756 CCATAAATGAAGTCTCTTTCCCCACCATTTGAATCTGGGCTGGCACTGTGACTTGATTTG 1815

Qy        311 A-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAAC 369
Db       1816 ATCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCCAAGCCTAGGTTTCAAGAGGCC 1875

Qy        370 TTACA 374
Db       1876 TTATA 1880

RESULT 7
US-08-642-274D-38
; Sequence 38, Application US/08642274D
; Patent No. 6200749
; GENERAL INFORMATION:
;  APPLICANT:  Shiloh, Yosef
;  TITLE OF INVENTION:  MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
;  TITLE OF INVENTION:  SCREEN FOR A PARTIAL A-T PHENOTYPE
;  FILE REFERENCE:  229000033
;  CURRENT APPLICATION NUMBER: US/08/642,274D
;  CURRENT FILING DATE:  1996-05-03
;  NUMBER OF SEQ ID NOS:  220
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 38
;   LENGTH: 537
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: genomic
US-08-642-274D-38

  Query Match             9.5%;  Score 47.8;  DB 3;  Length 537;
  Best Local Similarity  70.3%;  Pred. No. 2.3e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db         12 CCAAAGTGCTGGGATTACAGTCGTGAGCCACCGCACTCGGCCTTAAGGTTAATTCTTGAA 71

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAAT 91
Db         72 GTACAGAAAAACAGCATTATAGTTTGGAAAT 102

RESULT 8
US-08-952-014C-38
; Sequence 38, Application US/08952014C
; Patent No. 6265158
;  GENERAL INFORMATION:
;    APPLICANT:  Shiloh, Yosef
;    TITLE OF INVENTION:  ATAXIA-TELANGIECTASIA GENE AND ITS
;    TITLE OF INVENTION:  GENOMIC ORGANIZATION
;    NUMBER OF SEQUENCES:  91
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Kohn & Associates
;      STREET:  30500 No. 6265158thwestern Hwy., Suite 410
;      CITY:  Farmington Hills
;      STATE:  Michigan
;      COUNTRY:  U.S.
;      ZIP:  48334
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/952,014C
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kohn, Kenneth I.
;      REGISTRATION NUMBER:  30,995
;      REFERENCE/DOCKET NUMBER:  2290.00028
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  810-539-5050
;      TELEFAX:  810-539-5055
;  INFORMATION FOR SEQ ID NO:  38:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  537 base pairs
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-952-014C-38

  Query Match              9.5%;  Score 47.8;  DB 3;  Length 537;
  Best Local Similarity   70.3%;  Pred. No. 2.3e-05;
  Matches   64;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db         12 CCAAAGTGCTGGGATTACAGTCGTGAGCCACCGCACTCGGCCTTAAGGTTAATTCTTGAA 71

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAAT 91
Db         72 GTACAGAAAAACAGCATTATAGTTTGGAAAT 102


RESULT 9
US-09-499-522-5/c
; Sequence 5, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
;  APPLICANT: Blumenfeld, Marta
;  APPLICANT: Bougueleret, Lydie
;  APPLICANT: Bihain, Bernard
;  TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
;  FILE REFERENCE: GENSET.053AUS
;  CURRENT APPLICATION NUMBER: US/09/499,522
;  CURRENT FILING DATE: 2000-02-10
;  EARLIER APPLICATION NUMBER: US 60/119,592
;  EARLIER FILING DATE: 1999-02-10
;  EARLIER APPLICATION NUMBER: US 60/144,784
;  EARLIER FILING DATE: 1999-07-20
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: Patent.pm
; SEQ ID NO 5
;   LENGTH: 465
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 126
;   OTHER INFORMATION: 99-14417-126  :  polymorphic  base  C  or  T
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 103..125
;   OTHER INFORMATION: 99-14417-126.mis1  potential
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 127..149
;   OTHER INFORMATION: 99-14417-126.mis2  potential,  complement
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 334
;   OTHER INFORMATION: 99-14417-334  :  polymorphic  base  C  or  T
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 311..333
;   OTHER INFORMATION: 99-14417-334.mis1  potential
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 335..357
;   OTHER INFORMATION: 99-14417-334.mis2  potential
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 1..21
;   OTHER INFORMATION: upstream  amplification  primer
;   FEATURE:
;   NAME/KEY: misc_binding
;   LOCATION: 447..465
;   OTHER INFORMATION: downstream  amplification  primer,  complement
US-09-499-522-5

  Query Match              9.3%;  Score 46.6;  DB 4;  Length 465;
  Best Local Similarity   59.6%;  Pred. No. 5e-05;
  Matches  133;  Conservative    0;  Mismatches   84;  Indels    6;  Gaps    3;

Qy        223 TTGCAATGTGACTTTGCTACT-TCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTG 281
Db        434 TTACAAGGTGACTTTGCCACTCTTCCCATCAAGGGATGGAGTCTGTGTCTCCACCCCTTG 375

Qy        282 CACTAGAGTTGGCCTTCTGACTTGCTTT-GACAATGGAATGTAGTACAAATGACACTGTG 340
Db        374 AATCTGGGCTGACCTTGTAATTTCTTTTGGCCAATCAAACATGGTGAAAGTAATGCTGTG 315

Qy        341 CAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAAC 400
Db        314 TTAGTTTCAAGCCTAGACTTCAAGAAACCTTGCCGCTTCCATGTACTCTCTTAGTTGACC 255

Qy        401 ----CAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAG 439
Db        254 NTTGCCATGACCTTGTGAACATGACCGGGCTAGCCTGCTGGAG 212


RESULT 10
US-09-873-404-3
; Sequence 3, Application US/09873404
; Patent No. 6500656
; GENERAL INFORMATION:
;  APPLICANT: WEBSTER, Marion et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001212-CIP
;  CURRENT APPLICATION NUMBER: US/09/873,404
;  CURRENT FILING DATE:  2001-06-05
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 63588
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(63588)
;   OTHER INFORMATION: n = A,T,C or G
US-09-873-404-3

  Query Match              9.1%;  Score 45.6;  DB 4;  Length 63588;
  Best Local Similarity   54.9%;  Pred. No. 0.00076;
  Matches   90;  Conservative    0;  Mismatches   74;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
Db       4171 CCAAAGTGCTGGGATTACAGGCATGAGCCACCACACCCATCCAACATCTATGTTATTAAT 4230

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCC 120
Db       4231 CTATTGCTGTGTAGCATATTACCCCAAACTTAGTGGCTTAAAGAATAAACATTTATTGTC 4290

Qy        121 GTCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTG 164
Db       4291 TCAGAGATCCTGTGCATCAAGAATTTAGGATGACGATCATTGAG 4334


RESULT 11
US-09-345-882-1/c
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
;  APPLICANT: Bougueleret, Lydie
;  TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-
;  TITLE OF INVENTION:  AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
;  FILE REFERENCE: GENSET.031A
;  CURRENT APPLICATION NUMBER: US/09/345,882
```

```
;  CURRENT FILING DATE:  1999-06-30
;  PRIOR APPLICATION NUMBER: US 60/091,315
;  PRIOR FILING DATE: 1998-06-30
;  PRIOR APPLICATION NUMBER: US 60/111,909
;  PRIOR FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 140
;  SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 162480
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 72794
;   OTHER INFORMATION: 5-124-273  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 88075
;   OTHER INFORMATION: 5-127-261  :  polymorphic  base  A  or  C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 90842
;   OTHER INFORMATION: 99-1437-325  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 93714
;   OTHER INFORMATION: 5-128-60  :  polymorphic  base  deletion  of  GT
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97122
;   OTHER INFORMATION: 99-1442-224  :  polymorphic  base  G  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97152
;   OTHER INFORMATION: 5-129-144  :  polymorphic  base  deletion  of  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99098
;   OTHER INFORMATION: 5-130-257  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99117
;   OTHER INFORMATION: 5-130-276  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 103806
;   OTHER INFORMATION: 5-131-395  :  polymorphic  base  A  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 106940
;   OTHER INFORMATION: 5-133-375  :  polymorphic  base  insertion  of  A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108106
;   OTHER INFORMATION: 5-135-155  :  polymorphic  base  insertion  of  A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108149
;   OTHER INFORMATION: 5-135-198  :  polymorphic  base  insertion  of  GTTT
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108308
;   OTHER INFORMATION: 5-135-357  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108471
;   OTHER INFORMATION: 5-136-174  :  polymorphic  base  C  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 134134
;   OTHER INFORMATION: 5-140-120  :  polymorphic  base  C  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 134362
;   OTHER INFORMATION: 5-140-348  :  polymorphic  base  insertion  of  A
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 134374
;   OTHER INFORMATION: 5-140-361  :  polymorphic  base  insertion  of  CA
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 146328
;   OTHER INFORMATION: 5-143-84  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 146345
;   OTHER INFORMATION: 5-143-101  :  polymorphic  base  A  or  C
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 150329
;   OTHER INFORMATION: 5-145-24  :  polymorphic  base  A  or  G
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 160031
;   OTHER INFORMATION: 5-148-352  :  polymorphic  base  G  or  T
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 72771..72817
;   OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 72771..72817
;   OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 88050..88096
;   OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 88050..88096
;   OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 90819..90865
;   OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 90819..90865
;   OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 93690..93736
;   OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 93690..93736
;   OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97099..97145
;   OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97099..97145
;   OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97130..97177
;   OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 97130..97177
;   OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99075..99121
```

```
;   OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID34
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99075..99121
;   OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID55
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99094..99140
;   OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID35
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 99094..99140
;   OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID56
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 103783..103828
;   OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID36
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 103783..103828
;   OTHER INFORMATION: polymorphic fragment 5-131-395 SEQ ID57
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 106918..106966
;   OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID37
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 106918..106966
;   OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID58
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108084..108130
;   OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID38
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108084..108130
;   OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID59
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108127..108177
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID39
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 108127..108177
;   OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID60
;   FEATURE:

  Query Match             9.0%;  Score 45.2;  DB 4;  Length 162450;
  Best Local Similarity   75.7%;  Pred. No. 0.0015;
  Matches   56;  Conservative    0;  Mismatches   18;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
              ||||| | ||||| ||||  |||  |||  |    ||   |   |    |
Db     101563 CCAAAGTGCTGGGATTACAGGCGTGAGCCACCGCGCCCAGCCACACCCAGCTAATTTTTG 101504

Qy         61 TTTTTATAATTCAG 74
               ||  ||  |||
Db     101503 TATTTTTAGTACAG 101490


RESULT 12
US-09-705-299-11/c
; Sequence 11, Application US/09705299
; Patent No. 6440737
; GENERAL INFORMATION:
;  APPLICANT:  Lex M. Cowsert
;  APPLICANT:  Susan M. Freier
;  TITLE OF INVENTION:  ANTISENSE MODULATION OF CELLULAR APOPTOSIS SUSCEPTIBILITY GENE 5
;  FILE REFERENCE:  RTS-0174
;  CURRENT APPLICATION NUMBER: US/09/705,299
;  CURRENT FILING DATE:  2000-11-01
;  NUMBER OF SEQ ID NOS:  86
; SEQ ID NO 11
;   LENGTH: 3609
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: 92
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 569
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 1385
;   OTHER INFORMATION: unknown
US-09-705-299-11

  Query Match             9.0%;  Score 45;  DB 4;  Length 3609;
  Best Local Similarity   65.3%;  Pred. No. 0.00036;
  Matches   66;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
              |||| |||||||||||||||| || |||||   ||||  |||  ||| |    ||
Db       2024 CCAAAGTGCTGGGATTACAGGTGTGAGCCACCGCATCCGGCCTAAAGCAGGCATGTTTTA 1965

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAA 101
                 |   | |     |       | |  | ||  || |
Db       1964 AGTGGGTATTTTGAAAAGGTCAAAGAACAAAGCATTCTAAA 1924


RESULT 13
US-09-735-934A-3
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:
;  APPLICANT: LI, Jiayin et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL000851
;  CURRENT APPLICATION NUMBER: US/09/735,934A
;  CURRENT FILING DATE:  2000-12-14
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 43950
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-735-934A-3

  Query Match             8.9%;  Score 44.4;  DB 4;  Length 43950;
  Best Local Similarity   77.1%;  Pred. No. 0.0018;
  Matches   54;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60
               ||| |||  ||| ||||     |||   |||| | |    ||||     ||
Db      22734 CCAAAGTGCTGGGAGTACAGGCACGAGCCACCACATCCGGCCATCAAAATGTTTATCAAG 22793

Qy         61 TTTTTATAAT 70
                |||  ||
Db      22794 CTTTTACTAT 22803


RESULT 14
US-10-060-332-3
; Sequence 3, Application US/10060332
; Patent No. 6528294
; GENERAL INFORMATION:
;  APPLICANT: LI, Jiayin et al
;  TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL000851DIV
;  CURRENT APPLICATION NUMBER: US/10/060,332
```

```
;  CURRENT FILING DATE:  2002-02-01
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 43950
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-060-332-3

  Query Match             8.9%;  Score 44.4;  DB 4;  Length 43950;
  Best Local Similarity   77.1%;  Pred. No. 0.0015;
  Matches   54;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60

Db      22734 CCAAAGTGCTGGGAGTACAGGCACGAGCCACCACATCCGGCCATCAAAATGTTTATCAAG 22793

Qy         61 TTTTTATAAT 70

Db      22794 CTTTTACTAT 22803


RESULT 15
US-08-076-011-1
; Sequence 1, Application US/08076011
; Patent No. 5521069
;  GENERAL INFORMATION:
;    APPLICANT:  ONDA, Haruo
;    APPLICANT:  KIMURA, Chiharu
;    APPLICANT:  OHKUBO, Shoichi
;    TITLE OF INVENTION:  NOVEL DNA AND USE THEREOF
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DAVID G. CONLIN; DIKE, BRONSTEIN,
;      ADDRESSEE:  ROBERTS & CUSHMAN
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/076,011
;      FILING DATE:  11-JUN-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/047,246
;      FILING DATE:  13-APR-1993
;      NAME:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/741,676
;      FILING DATE:  07-AUG-1991
;      NAME:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RESNICK, DAVID S
;      REGISTRATION NUMBER:  34235
;      REFERENCE/DOCKET NUMBER:  41155-CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)523-3400
;      TELEFAX:  (617)523-6440
;      TELEX:  200291 STRE UR
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  17041 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  join(7540..7650, 9814..9945, 10421..10519,
;      LOCATION:  11602..11787)
US-08-076-011-1

  Query Match             8.8%;  Score 44;  DB 1;  Length 17041;
  Best Local Similarity   63.0%;  Pred. No. 0.0024;
  Matches   68;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy          1 CCAAAGTGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAG 60

Db      14792 CCAAAGTGCTGGGATTACAGGCGTGAGCCACCACACCCAGCCTAAAATAGGGATCTAACA 14851

Qy         61 TTTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGAT 108

Db      14852 TTGTTCTTATACAAGTAACTCTGCAGACTAAACTTGTCTTGATAAAAT 14899

Search completed: October 24, 2003, 15:42:46
Job time : 34.3542 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 133.415 Seconds
(without alignments)
10586.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaaagtgctgggattacag.....tgaatgaggctagctaggcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	68.8	13.7	756	13	US-10-027-632-18943	Sequence 18943, A
2	68.8	13.7	756	13	US-10-027-632-149822	Sequence 149822,
3	64.8	12.9	586	12	US-10-029-386-12238	Sequence 12238, A
c 4	64.6	12.9	5239	12	US-10-311-455-246	Sequence 246, App
c 5	64.6	12.9	3673778	12	US-10-312-841-2	Sequence 2, Appli
6	61	12.2	670	9	US-09-811-284-93	Sequence 93, Appl
7	60.6	12.1	624	13	US-10-027-632-55107	Sequence 55107, A
8	60.6	12.1	624	13	US-10-027-632-311174	Sequence 311174,
9	60.6	12.1	627	13	US-10-027-632-908	Sequence 908, App
10	60.4	12.1	5239	12	US-10-311-455-245	Sequence 245, App
11	60.4	12.1	3673778	12	US-10-312-841-1	Sequence 1, Appli
c 12	60.2	12.0	651	13	US-10-040-739-325	Sequence 925, App
c 13	59.4	11.9	594	13	US-10-027-632-67207	Sequence 67207, A
c 14	56.6	11.3	646	13	US-10-027-632-16407	Sequence 16407, A
c 15	55	11.0	546	13	US-10-027-632-160975	Sequence 160975,
c 16	55	11.0	546	13	US-10-027-632-160976	Sequence 160976,
c 17	55	11.0	546	13	US-10-027-632-160977	Sequence 160977,
c 18	54.6	10.9	442	13	US-10-027-632-318161	Sequence 318161,
c 19	54.6	10.9	450	13	US-10-027-632-93145	Sequence 93145, A
20	53.8	10.7	590	13	US-10-027-632-85250	Sequence 85250, A
21	53.8	10.7	590	13	US-10-027-632-314433	Sequence 314433,
22	53.8	10.7	604	13	US-10-027-632-105017	Sequence 105017,
23	53.8	10.7	624	13	US-10-027-632-214356	Sequence 214356,
c 24	53.8	10.7	706	13	US-10-027-632-13713	Sequence 13713, A
25	53.6	10.7	1586	9	US-09-870-962-11	Sequence 11, Appl
26	53.6	10.7	1868	14	US-10-121-925-1	Sequence 1, Appli
27	53.6	10.7	2142	14	US-10-165-696-3	Sequence 3, Appli
c 28	53.4	10.7	584	13	US-10-027-632-46665	Sequence 46665, A
c 29	53.4	10.7	584	13	US-10-027-632-295733	Sequence 295733,
c 30	53.2	10.6	716	13	US-10-027-632-270803	Sequence 270803,
c 31	53.2	10.6	716	13	US-10-027-632-270804	Sequence 270804,
32	53.2	10.6	752	13	US-10-027-632-174127	Sequence 174127,
c 33	53.2	10.6	819	13	US-10-027-632-138083	Sequence 138083,
34	52.8	10.5	556	13	US-10-027-632-227143	Sequence 227143,
c 35	52.6	10.5	397658	10	US-09-813-320-3	Sequence 3, Appli
36	52.2	10.4	2141	14	US-10-165-696-1	Sequence 1, Appli
c 37	52	10.4	857	13	US-10-027-632-167271	Sequence 167271,
c 38	52	10.4	857	13	US-10-027-632-167272	Sequence 167272,
c 39	52	10.4	857	13	US-10-027-632-167273	Sequence 167273,
c 40	52	10.4	857	13	US-10-027-632-167274	Sequence 167274,
41	51.6	10.3	1434	10	US-09-764-855-163	Sequence 163, App
42	51.6	10.3	1434	14	US-10-072-349-163	Sequence 163, App
c 43	50.6	10.1	702	13	US-10-027-632-137333	Sequence 137333,
c 44	50.4	10.1	522	13	US-10-027-632-51013	Sequence 51013, A
45	50.4	10.1	522	13	US-10-027-632-85803	Sequence 85803, A

ALIGNMENTS

RESULT 1
US-10-027-632-18943
; Sequence 18943, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18943
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-18943

Query Match 13.7%; Score 68.8; DB 13; Length 756;
Best Local Similarity 62.9%; Pred. No. 1e-09;
Matches 140; Conservative 0; Mismatches 82; Indels 2; Gaps 2;

Qy 156 ATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATA 215
Db 109 ATCATTATGGTGGAGTCTTTGAAAAAGTGGCTACAACAAATTCTCCCATCCCTGTGTGCA 168

```
Qy        216 TGCCCCTTTGCAATGTGACTTTGCTACTTCT-CTATCAAGATGTGGAGCTTATTTTCCCA 274

Db        169 TACCGCTTTCCAAAGTGCCTTTGTAGCTCCTCCTGTCAAGAGGTAGGATCTGTTTTTCTA 228

Qy        275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 333

Db        229 CCCATACAATCTAGACTGGCCATATGATTTCCTTTGACCAAGAGAATGTTCTGGCAATGA 288

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCT 377

Db        289 CCCTGTGGACCTCTAAACCTAGACTTGAAGAACCATTTCACCT 332
```

RESULT 2
US-10-027-632-149822
; Sequence 149822, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 149822
; LENGTH: 756
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-149822

```
  Query Match             13.7%;  Score 68.8;  DB 13;  Length 756;
  Best Local Similarity   62.5%;  Pred. No. 1e-09;
  Matches  140;  Conservative    0;  Mismatches   82;  Indels    2;  Gaps    2;

Qy        156 ATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATA 215

Db        109 ATCATTATGGTGGAGTCTTTGAAAAAGTGGCTACAACAAATTCTCCCATCCCTGTGTGCA 168

Qy        216 TGCCCCTTTGCAATGTGACTTTGCTACTTCT-CTATCAAGATGTGGAGCTTATTTTCCCA 274

Db        169 TACCGCTTTCCAAAGTGCCTTTGTAGCTCCTCCTGTCAAGAGGTAGGATCTGTTTTTCTA 228

Qy        275 TATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGA 333

Db        229 CCCATACAATCTAGACTGGCCATATGATTTCCTTTGACCAAGAGAATGTTCTGGCAATGA 288

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCT 377

Db        289 CCCTGTGGACCTCTAAACCTAGACTTGAAGAACCATTTCACCT 332
```

RESULT 3
US-10-029-386-12238
; Sequence 12238, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12238
; LENGTH: 586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011977.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.46
; OTHER INFORMATION: NT HIT: gi9626821, EVALUE 1.00e-04
; OTHER INFORMATION: SWISSPROT HIT: P31554, EVALUE 7.40e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE067351.1, EVALUE 2.00e-04
US-10-029-386-12238

```
  Query Match             12.9%;  Score 64.8;  DB 12;  Length 586;
  Best Local Similarity   63.0%;  Pred. No. 1.3e-08;
  Matches  116;  Conservative    0;  Mismatches   67;  Indels    1;  Gaps    1;

Qy        170 TTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAAT 229

Db        356 TCATTTGCAAAGATGGTCTCAATAATTTCTCCCATGTCTATGGGCAGGTCTTTTTGCAGT 415

Qy        230 GTGACTTTGCTACT-TCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGA 288

Db        416 GTGACTTTGCCACTCTTTCTGTCAAGAAGTAGAGTTTATTTTCCATCCTCTTGCATCTAG 475

Qy        289 GTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTG 348

Db        476 TCTGGTCTTGTACTTGTTTTATCAAGTAAAATGTGGTAGAAGTGACTAGGGGATTCCTGA 535

Qy        349 GATT 352

Db        536 GCTT 539
```

RESULT 4
US-10-311-455-246/c
; Sequence 246, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 246
; LENGTH: 5239
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-246

```
  Query Match             12.9%;  Score 64.6;  DB 12;  Length 5239;
  Best Local Similarity   51.3%;  Pred. No. 4.8e-08;
  Matches  174;  Conservative    0;  Mismatches  164;  Indels    1;  Gaps    1;
```

```
Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db       3273 TTTAATAACTTTAAAAAACACCCAAAATTTAATAATTTAAATATTCAATCAAATTAAACT 3214

Qy        122 TCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db       3213 TCATTACGTTCTAATAACCAACAACCTAAAAATATTTATTACCATAAAAAAACTTACAAA 3154

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db       3153 ATATAACCATCAATCTCTCCTATTCTAATATACCTACCCCTTTACCGTATAACTTTACCA 3094

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db       3093 TTCCTCCTATCAAAAAATAAATTCTATACCTCCAATCTTAAATCTAAACTAACCTTATAA 3034

Qy        302 CTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       3033 TTTACTTTAACCAATAAAATATAACAAAAATAATATTATATAACTTTTAAAACTAAACCT 2974

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAA 399
Db       2973 CGAAAAACCTTACAACTTATATTTTAAATTCCTCAAAAA 2935


RESULT 5
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
;  APPLICANT: Epigenomics AG
;  TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
;  FILE REFERENCE: E01/1208/WO
;  CURRENT APPLICATION NUMBER: US/10/312,841
;  CURRENT FILING DATE:  2002-12-30
;  NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
;  LENGTH: 3673778
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
;  FEATURE:
;  NAME/KEY: unsure
;  LOCATION: (3379615)
US-10-312-841-2

  Query Match           12.9%;  Score 64.6;  DB 12;  Length 3673778;
  Best Local Similarity   51.3%;  Pred. No. 1.5e-06;
  Matches  174;  Conservative    0;  Mismatches  164;  Indels    1;  Gaps    1;

Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db    1463204 TTTAATAACTTTAAAAAACACCCAAAATTTAATAATTTAAATATTCAATCAAATTAAACT 1463145

Qy        122 TCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db    1463144 TCATTACGTTCTAATAACCAACAACCTAAAAATATTTATTACCATAAAAAAACTTACAAA 1463085

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db    1463084 ATATAACCATCAATCTCTCCTATTCTAATATACCTACCCCTTTACCGTATAACTTTACCA 1463025

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db    1463024 TTCCTCCTATCAAAAAATAAATTCTATACCTCCAATCTTAAATCTAAACTAACCTTATAA 1462965

Qy        302 CTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db    1462964 TTTACTTTAACCAATAAAATATAACAAAAATAATATTATATAACTTTTAAAACTAAACCT 1462905

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAA 399
Db    1462904 CGAAAAACCTTACAACTTATATTTTAAATTCCTCAAAAA 1462866


RESULT 6
US-09-811-284-93
; Sequence 93, Application US/09811284
; Patent No. US20020058306A1
; GENERAL INFORMATION:
;  APPLICANT: Vogeli, Gabriel
;  TITLE OF INVENTION: No. US20020058306A1el G Protein-Coupled Receptors
;  FILE REFERENCE: 00167US1
;  CURRENT APPLICATION NUMBER: US/09/811,284
;  CURRENT FILING DATE:  2001-03-16
;  PRIOR APPLICATION NUMBER: 60/189,783
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/189,907
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/189,918
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/189,960
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/189,917
;  PRIOR FILING DATE: 2000-03-16
;  PRIOR APPLICATION NUMBER: 60/192,945
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 60/192,916
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 60/192,923
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 60/192,933
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 60/192,830
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: 60/192,234
;  PRIOR FILING DATE: 2000-03-27
;  PRIOR APPLICATION NUMBER: 60/192,155
;  PRIOR FILING DATE: 2000-03-24
;  PRIOR APPLICATION NUMBER: 60/192,935
;  PRIOR FILING DATE: 2000-03-29
;  NUMBER OF SEQ ID NOS: 258
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 93
;  LENGTH: 670
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-811-284-93

  Query Match           12.2%;  Score 61;  DB 9;  Length 670;
  Best Local Similarity   60.6%;  Pred. No. 1.8e-07;
  Matches  117;  Conservative    0;  Mismatches   75;  Indels    1;  Gaps    1;

Qy        237 TGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCT 296
Db        319 TGCAGTGTTTTCATCAAGAGCTGCAGTCTATTTCCCCACCCCAAAATCCA-AGGTGGCCT 377

Qy        297 TCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAG 356
Db        378 TGTGACTTGCTTTGAAAATGAAATGCAGTGGAAGGAATTTTGTGTGACTTTCCAGACTAG 437

Qy        357 GTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAG 416
Db        438 GCTGCAAGAGACTTTGCAGATTCCCCTTCCACATTCTTGGAATGCTGCCCTGAGATAGCC 497

Qy        417 AAGTCAGGGCTAT 429
Db        498 ATGCAAGGAAGAT 510


RESULT 7
US-10-027-632-55107
; Sequence 55107, Application US/10027632
; GENERAL INFORMATION:
```

```
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 55107
;   LENGTH: 624
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-55107

  Query Match             12.1%;  Score 60.6;  DB 13;  Length 624;
  Best Local Similarity   61.7%;  Pred. No. 2.3e-07;
  Matches  113;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy        179 AAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTG 238

Db        442 AAAAATACCTCCATGTCTCCACCCTTCCTGGTATACACACTACATTATAATATGGCTTTG 501

Qy        239 CTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTC 298

Db        502 CTGCTCTTCTCATCAAGAGTGGAATCGACTTACCTAATCCTTGAATGTGAGCTGGCCTTG 561

Qy        299 TGACTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGG 357

Db        562 TAACTTGCTTTGACCAATGGAATGTGCCAGAAGCGACCTTGTGCCAATTCTGAGTCTAAG 621

Qy        358 TTT 360

Db        622 CTT 624

RESULT 8
US-10-027-632-311174
; Sequence 311174, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 311174
;   LENGTH: 624
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-311174

  Query Match             12.1%;  Score 60.6;  DB 13;  Length 624;
  Best Local Similarity   61.7%;  Pred. No. 2.3e-07;
  Matches  113;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy        179 AAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTG 238

Db        442 AAAAATACCTCCATGTCTCCACCCTTCCTGGTATACACACTACATTATAATATGGCTTTG 501

Qy        239 CTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTC 298

Db        502 CTGCTCTTCTCATCAAGAGTGGAATCGACTTACCTAATCCTTGAATGTGAGCTGGCCTTG 561

Qy        299 TGACTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGG 357

Db        562 TAACTTGCTTTGACCAATGGAATGTGCCAGAAGCGACCTTGTGCCAATTCTGAGTCTAAG 621

Qy        358 TTT 360

Db        622 CTT 624

RESULT 9
US-10-027-632-908
; Sequence 908, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 908
;   LENGTH: 627
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-908

  Query Match             12.1%;  Score 60.6;  DB 13;  Length 627;
  Best Local Similarity   61.7%;  Pred. No. 2.3e-07;
  Matches  113;  Conservative    0;  Mismatches   69;  Indels    1;  Gaps    1;

Qy        179 AAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTG 238

Db        444 AAAAATACCTCCATGTCTCCACCCTTCCTGGTATACACACTACATTATAATATGGCTTTG 503

Qy        239 CTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTC 298

Db        504 CTGCTCTTCTCATCAAGAGTGGAATCGACTTACCTAATCCTTGAATGTGAGCTGGCCTTG 563
```

```
Qy        299 TGACTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTAGG 357
Db        564 TAACTTGCTTTGACCAATGGAATGTGCCAGAAGCGACCTTGTGCCAATTCTGAGTCTAAG 623

Qy        358 TTT 360
Db        624 CTT 626
```

RESULT 10
US-10-311-455-245
; Sequence 245, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 245
; LENGTH: 5239
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-245

```
  Query Match             12.1%;  Score 60.4;  DB 12;  Length 5239;
  Best Local Similarity   50.6%;  Pred. No. 8e-07;
  Matches  171;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db       1967 TTTGATGGTTTTAAAAGGTATTTAAAGTTTAGTGATTTGAATGTTTAGTTAGGTTAGGTT 2026

Qy        122 TCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db       2027 TTATTACGTTTTGGTAATTAATAATTTAGAAATATTTGTTGTTATAGGAGGGTTTATAAA 2086

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db       2087 ATATAGTTATTAATTTTTTTTATTTTGATGTGTTTGTTTTTTTGTCGTGTGATTTTGTTA 2146

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db       2147 TTTTTTTTATTAAGAGGTAAATTTTATGTTTTTAGTTTTAAATTTGGGTTGATTTTGTGA 2206

Qy        302 CTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db       2207 TTTGTTTTGATTAATAGAATGTGGTAGAAGTGATGTTATGTGATTTTTGGGGTTAGGTTT 2266

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAA 398
Db       2267 CGAGAGATTTTGTAGTTTATGTTTTGGATTTTTTAGAA 2304
```

RESULT 11
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

```
  Query Match             12.1%;  Score 60.4;  DB 12;  Length 3673778;
  Best Local Similarity   50.6%;  Pred. No. 2.4e-05;
  Matches  171;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy         62 TTTTATAATTCAGCAAAATGGTTGGAAAATGCTGTCTTAAATGAGATGCTTAAGCTGCCG 121
Db    2210575 TTTGATGGTTTTAAAAGGTATTTAAAGTTTAGTGATTTGAATGTTTAGTTAGGTTAGGTT 2210634

Qy        122 TCTGAACATGAGGTAGAAGGAAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAG 181
Db    2210635 TTATTACGTTTTGGTAATTAATAATTTAGAAATATTTGTTGTTATAGGAGGGTTTATAAA 2210694

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241
Db    2210695 ATATAGTTATTAATTTTTTTTATTTTGATGTGTTTGTTTTTTTGTCGTGTGATTTTGTTA 2210754

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301
Db    2210755 TTTTTTTTATTAAGAGGTAAATTTTATGTTTTTAGTTTTAAATTTGGGTTGATTTTGTGA 2210814

Qy        302 CTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360
Db    2210815 TTTGTTTTGATTAATAGAATGTGGTAGAAGTGATGTTATGTGATTTTTGGGGTTAGGTTT 2210874

Qy        361 CGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAA 398
Db    2210875 CGAGAGATTTTGTAGTTTATGTTTTGGATTTTTTAGAA 2210912
```

RESULT 12
US-10-040-739-925/c
; Sequence 925, Application US/10040739
; Publication No. US20020173635A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; McCoy, John
; LaVallie, Edward
; Racie, Lisa
; Merberg, David
; Treacy, Maurice
; Spaulding, Vikki
; TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
; NUMBER OF SEQUENCES: 1519
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/040,739
; FILING DATE: 07-Jan-2002

```
;           CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: 09/036,520
;           FILING DATE: 03-JUN-1998
;        ATTORNEY/AGENT INFORMATION:
;           NAME: Brown, Scott A.
;           REGISTRATION NUMBER: 32,724
;        TELECOMMUNICATION INFORMATION:
;           TELEPHONE: (617) 498-8224
;           TELEFAX: (617) 876-5851
;   INFORMATION FOR SEQ ID NO: 925:
;        SEQUENCE CHARACTERISTICS:
;           LENGTH: 651 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: double
;           TOPOLOGY: linear
;        MOLECULE TYPE: cDNA
;        SEQUENCE DESCRIPTION: SEQ ID NO: 925:
US-10-040-739-925

  Query Match             12.0%;  Score 60.2;  DB 13;  Length 651;
  Best Local Similarity   61.6%;  Pred. No. 3.1e-07;
  Matches  146;  Conservative    0;  Mismatches   88;  Indels    3;  Gaps    3;

Qy        182 ATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTA 241

Db        368 AAGGCCACAAATGAGTAACTCCTTCCCATGTCCAGACTCCTTTGCAATGTGACCTTGCTA 309

Qy        242 CTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGA 301

Db        308 CTTCTC-ACCAGGAAATTGAGTCTATTTTTCCATTCTTTTGAATCTGGCTGGCCTTGTGA 250

Qy        302 CTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTT-TGGATTTTAGGTT 359

Db        249 CTTGCTTTGACTAACACAAGGTGGCAGAAGTATCACTGTATGAGTTCTTGGGCATAGGGC 190

Qy        360 TCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATGTAAAG 416

Db        189 TTAAGAGGTATTGCAAGTTCTACTCTTTTCTTTTTGGGACCCTAAGCCATCATGATG 133


RESULT 13
US-10-027-632-67207/c
; Sequence 67207, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 67207
;   LENGTH: 594
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-67207

  Query Match             11.9%;  Score 59.4;  DB 13;  Length 594;
  Best Local Similarity   63.3%;  Pred. No. 5.1e-07;
  Matches  107;  Conservative    0;  Mismatches   61;  Indels    1;  Gaps    1;

Qy        157 TCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATAT 216

Db        324 TTAGTGTGGTCAATTGTTTGCAAGGATGACTACAACACTTCTTGCTATCTCTGTGTGCGT 265

Qy        217 GCCCCTTTGCAATGTGACTTTGCTACTTCT-CTATCAAGATGTGGAGCTTATTTTCCCAT 275

Db        264 GCCCCCTTGCCGGGTGACTTTGCCAGTCCTCCCATGGTACAGTGGCATCTGTCTTCCCTT 205

Qy        276 ATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAG 324

Db        204 CATCTGTGCCTGGGCTAGGCCTCTGACTTTCTTTGACTATCAAAACTTG 156


RESULT 14
US-10-027-632-16407/c
; Sequence 16407, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16407
;   LENGTH: 646
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-16407

  Query Match             11.3%;  Score 56.6;  DB 13;  Length 646;
  Best Local Similarity   69.4%;  Pred. No. 3.5e-06;
  Matches   77;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy        383 TCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACA 442

Db        451 TCTCTTCCTCTTGGAAGCCAGATGCCATTTGAAAAAGACTGGGTCATCCTGCTGAAGACA 392

Qy        443 TATGTCCCAGCTAATAGCCACAATCAACCTCTGAACATATGAATGAGGCTA 493

Db        391 TGTGGCATAGATTATAGCCAACACCAATTCCCATACCTGTGAGTGAGGCCA 341


RESULT 15
US-10-027-632-160975/c
; Sequence 160975, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
```

```
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 160975
;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-160975

  Query Match             11.0%;  Score 55;  DB 13;  Length 546;
  Best Local Similarity   62.1%;  Pred. No. 9.3e-06;
  Matches  100;  Conservative    3;  Mismatches   53;  Indels    5;  Gaps    1;

Qy        206 CTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTT 265
              |    ||  ||  | ||||   |||   |||      |   ||     |    |
Db        283 CCCTGATCAATACCCCTTTGCAATGTGAYTTTGTAGGTYCTCTAATCAAATAT-----CT 229

Qy        266 ATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGT 325
              ||||        | | | |     ||   ||    ||    |||  |||    |    |
Db        228 ATTTTCMTACCCTTTGAATTTGGGCTGGTCTTGTCACTTCTTTTGACAATAAAATGCAAT 169

Qy        326 ACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAG 366
               || |||    |  |  ||     |   |||    |
Db        168 AGAAGTGATAGTGAACCAGTTCCAAGTCTGGGTCTTAACAG 128

Search completed: October 24, 2003, 18:34:26
Job time : 134.415 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 1046.7 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaaagtgctgggattacag..........tgaatgaggctagctaggcc 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781332 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	85	17.0	726	29	AG030296	AG030296 Pan trogl
c	2	77	15.4	670	29	BZ610078	BZ610078 WHADV23TF
	3	76.6	15.3	657	14	CB215585	CB215585 NISC_np07
c	4	73.8	14.7	442	28	B80948	B80948 CIT-HSP-202
	5	71	14.2	655	29	AG046887	AG046887 Pan trogl
	6	70.6	14.1	608	28	AQ351210	AQ351210 RPCI11-12
	7	67.8	13.5	555	28	AQ614358	AQ614358 HS_5126_B
	8	67.4	13.5	722	13	BX103709	BX103709 BX103709
c	9	66.4	13.3	588	9	AL703889	AL703889 DKFZp686B
c	10	65.6	13.1	477	9	AI041146	AI041146 ov69f02.x
c	11	65	13.0	474	10	BF929327	BF929327 IL2-NT020
	12	64.8	12.9	401	28	AQ080797	AQ080797 CIT-HSP-2
	13	64.8	12.9	1335	11	AK046951	AK046951 Mus muscu
c	14	63.8	12.7	865	29	CC142611	CC142611 NDL.37D22
c	15	63.6	12.7	497	13	BX116065	BX116065 BX116065
	16	63.4	12.7	518	13	BU656781	BU656781 c116h09.z
	17	63	12.6	726	28	AQ389497	AQ389497 RPCI11-14
c	18	62.2	12.4	492	28	B39328	B39328 HS-1049-B1-
c	19	62	12.4	468	28	AQ239410	AQ239410 CIT-HSP-2
	20	62	12.4	540	28	AQ175020	AQ175020 HS_3210_A
	21	61.4	12.3	526	28	AQ175021	AQ175021 HS_3210_A
	22	59.6	11.9	457	28	AQ769797	AQ769797 HS_3170_B
	23	59.2	11.8	373	28	B83258	B83258 RPCI11-25K1
	24	59.2	11.8	498	28	AQ014114	AQ014114 RPCI11-25
c	25	59	11.8	502	2	HSM073396	Bx483228 Homo sapi
c	26	59	11.8	529	10	BF435562	BF435562 nac33a02.
	27	58.6	11.7	510	14	CB270506	CB270506 1009413 H
c	28	57.6	11.5	1201	9	AL579387	AL579387 AL579387
c	29	55.8	11.1	680	13	BU685340	BU685340 UI-CF-DU1
	30	55	11.0	511	28	AQ894910	AQ894910 HS_3193_A
c	31	54.2	10.8	421	28	AQ194975	AQ194975 RPCI11-57
c	32	53.6	10.7	453	9	AI334217	AI334217 qq09f12.x
c	33	53.6	10.7	453	9	AA582672	AA582672 nn53h12.s
	34	53.6	10.7	459	14	CB069503	CB069503 is15h09.y
	35	53.6	10.7	461	28	AQ003949	AQ003949 CIT-HSP-2
c	36	53.6	10.7	469	9	AI573185	AI573185 tr70f04.x
c	37	53.6	10.7	506	9	AA706818	AA706818 zj30e03.s
c	38	53.6	10.7	514	13	BU076998	BU076998 in20a05.x
c	39	53.6	10.7	556	14	CD369377	CD369377 UI-H-FT1-
c	40	53.6	10.7	622	14	CA427318	CA427318 UI-H-FH1-
c	41	53.6	10.7	633	14	CB528830	CB528830 UI-H-FT2-
c	42	53.6	10.7	666	14	CD370383	CD370383 UI-H-FT1-
c	43	53.6	10.7	670	13	BQ575071	BQ575071 UI-H-EZ1-
	44	53.6	10.7	697	13	BX106138	BX106138 BX106138
c	45	53.6	10.7	956	13	BX352121	BX352121 BX352121

ALIGNMENTS

RESULT 1
AG030296/c
LOCUS AG030296 726 bp DNA linear GSS 01-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-002K06.R, genomic survey sequence.
ACCESSION AG030296
VERSION AG030296.1 GI:16557169
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 726)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of

```
         clone tracking errors.
          PRIMERS
             Sequencing: M13Rev
          LIBRARY
             Vector     : pKS145
             R.Site 1   : SacI
             R.Site 2   : SacI.
FEATURES             Location/Qualifiers
     source          1..726
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-002K06.R"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      196 a    167 c    155 g    207 t      1 others
ORIGIN

  Query Match             17.0%;  Score 65;  DB 29;  Length 726;
  Best Local Similarity   62.5%;  Pred. No. 0.00034;
  Matches  168;  Conservative    0;  Mismatches   95;  Indels    6;  Gaps    2;

Qy        144 ATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTA 203

Db        437 AATTTGAAGATGTCCTATACGGTCGATTGTTTCCAAAGACGACTGCAACACTCCATTCTA 378

Qy        204 TCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGC 263

Db        377 TC-----TTACATGCCCTTTTGCAATATGACTTTTCCACTTTTCCATCAAGACATAGACT 323

Qy        264 TTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGT 322

Db        322 CTATCTTACCTCCCCTTAAATCTGGACTAATCTTATGAGTTGCTTTAACCAATATCATGT 263

Qy        323 AGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCAC 392

Db        262 GGTAGAAATGACAATGTGCAACTCCAAAGACAAAACTTCAAGAATCCCCACACCTTCCAT 203

Qy        383 TCACACTCTCTTGGAAACCAGATGCAATG 411

Db        202 TTTTTATGTCTTGGAATGCAGCTGCCATG 174


RESULT 2
BZ610078/c
LOCUS       BZ610078               670 bp    DNA     linear   GSS 08-JUN-2003
DEFINITION  WHADV23TF Human MCF7 breast cancer cell line library (MCF7_1) Homo
            sapiens genomic clone MCF7_1-24C22, genomic survey sequence.
ACCESSION   BZ610078
VERSION     BZ610078.1  GI:31518639
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Volik,S., Zhao,S., Chin,K., Brebner,J.H., Herndon,D.R., Tao,Q.,
            Kowbel,D., Huang,G., Lapuk,A., Kuo,W.-L., Magrane,G., de Jong,P.,
            Gray,J.W. and Collins,C.
  TITLE     ESP: a sequence-based approach to the structural genomics of tumors
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Volik SV
            Colin Collins' lab
            UCSF Comprehensive Cancer Center
            UCSF Box 0808, San Francisco, CA 94143-0808, USA
            Tel: 415 502 7066
            Fax: 415 502 5665
            Email: svolik@cc.ucsf.edu
            This clone is available from Amplicon Express
            http://www.genomex.com
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="MCF7_1-24C22"
                     /sex="female"
                     /clone_lib="Human MCF7 breast cancer cell line library
                     (MCF7_1)"
                     /note="Vector: pECBAC1; Site_1: HindIII; This library was
                     constructed from MCF7 breast cancer cell line by Amplicon
                     Express (http://www.genomex.com) using their standard
                     procedure."
BASE COUNT      215 a    159 c    118 g    178 t
ORIGIN

  Query Match             15.4%;  Score 77;  DB 29;  Length 670;
  Best Local Similarity   61.4%;  Pred. No. 0.0063;
  Matches  143;  Conservative    0;  Mismatches   85;  Indels    5;  Gaps    1;

Qy        154 TAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATA 213

Db        265 TAGGAAGTGTGGTAGATTGCTTGCAAAGATGACCAAAACAATTCCTTTTATCCTGC---- 210

Qy        214 TATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCC 273

Db        209 -ATGCCTTTTTGCAGTGTGATTTTGCCAATTTACTATTAAGAGGTGGAGTCTACTTCCCA 151

Qy        274 ATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGA 333

Db        150 ACCCCTTGAATCTGGGAGGGGCTTGTAATTTGCTTTGTCCAATAGAATGTATCGAAGGTG 91

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCAC 386

Db         90 ACCTCTGTGACTTCTGAGACCAGAGGTGAGGAGGCCTTGCAGCTTTCCCCCTC 38


RESULT 3
CB215585
LOCUS       CB215585               657 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  NISC_np07f11.y1 NICHD_HS_Ut1 Homo sapiens cDNA clone IMAGE:5937188
            5', mRNA sequence.
ACCESSION   CB215585
VERSION     CB215585.1  GI:28263777
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 657)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation:
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
             DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC)
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            info@image.llnl.gov
            Plate: LLAM13163  row: L  column: 21
            Seq primer: M13RP1 reverse primer (ABI).
FEATURES             Location/Qualifiers
     source          1..657
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5937188"
                     /sex="female"
                     /tissue_type="normal endometrium, late proliferative phase
                     , cycle day 13"
```

```
                    /lab_host="DH10B (T1-resistant)"
                    /clone_lib="NICHD_HS_Ut1"
                    /note="Organ: uterus; Vector: pCMV-SPORT6.1.ccdb (ResGen,
                    Invitrogen Corporation); Site_1: NotI; Site_2: EcoRV;
                    Cloned unidirectionally from microquantity amounts of mRNA
                    from normal endometrial tissue (late proliferative phase,
                    cycle day 13). Average insert size 1.9 kb. Library
                    constructed by ResGen (Invitrogen Corporation)."
BASE COUNT      196 a    156 c    125 g    180 t
ORIGIN

  Query Match              15.3%;  Score 76.6;  DB 14;  Length 657;
  Best Local Similarity    62.3%;  Pred. No. 0.0073;
  Matches  154;  Conservative    0;  Mismatches   89;  Indels    4;  Gaps    2;

Qy        196 TCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGA 255

Db          5 TCCACCCTTCCTGGTATACACACTACATTATAATATGGCTTTGCTGCTCTTCTCATCAAG 64

Qy        256 TGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGA-CAA 314

Db         65 AGTGGAATCGACTTACCTAATCCTTGAATGTGAGCTGGCCTTGTAACTTGCTTTGACCAA 124

Qy        315 TGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACA 374

Db        125 TGGAATGTGCCAGAAGCGACCTTGTGCCAATTCTGAGTCTAAGCTTCAAGAGGCCTTGTG 184

Qy        375 CCTTCCA---CTCACACTCTCTTGGAAACCAGATGCAATGTAAAGAAGTCAGGGCTATCC 431

Db        185 TGTTTCAACTCTCCCTCTTAGAGCCCTGCCAGATACCATGTGATGAAGCCCAGGCTAGCC 244

Qy        432 TGCTAGA 438

Db        245 TGGTAGA 251


RESULT 4
B80948/c
LOCUS       B80948                   442 bp    DNA     linear   GSS 25-JUN-1998
DEFINITION  CIT-HSP-2025N4.TFB CIT-HSP Homo sapiens genomic clone 2025N4,
            genomic survey sequence.
ACCESSION   B80948
VERSION     B80948.1  GI:2867971
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CIT-HSP-2025N4.TR CIT-HSP-2025N4.TF
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..442
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7047102"
                     /db_xref="taxon:9606"
                     /clone="2025N4"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      133 a    111 c     93 g    105 t
ORIGIN

  Query Match              14.7%;  Score 73.8;  DB 28;  Length 442;
  Best Local Similarity    60.5%;  Pred. No 0.02;
  Matches  141;  Conservative    0;  Mismatches   87;  Indels    5;  Gaps    1;

Qy        154 TAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATA 213

Db        319 TAGAAAGTGTGGTAGATTGCTTGCAAAGATGACCAAATCAATTCCTTTTATCCTGC---- 264

Qy        214 TATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCC 273

Db        263 -ATGCCTTTTTGCAGTGTGATTTTGCCAATTTACTATTAAGAGGTGGAGTCTACTTCCCA 205

Qy        274 ATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGA 333

Db        204 ACCCCTTGAATCTGGGAGGGGCTTGTAATTTGCTTTGTCCAATAGAATGTATCGAAGGTG 145

Qy        334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCAC 386

Db        144 ACCTCTGTGACTTCTGAGACCAGAGGTGAGGAGGCCATGCAGCTTTCCCCCTC 92


RESULT 5
AG046887
LOCUS       AG046887                 655 bp    DNA     linear   GSS 02-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-026D22.F, genomic survey sequence.
ACCESSION   AG046887
VERSION     AG046887.1  GI:16583779
KEYWORDS    GSS.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 655)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
             PRIMERS
                Sequencing: -21M13
             LIBRARY
                Vector      : pKS145
                R.Site 1    : SacI
                R.Site 2    : SacI.
FEATURES             Location/Qualifiers
     source          1..655
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="PTB-026D22.F"
                     /sex="male"
                     /cell_type="lymphoblast"
```

```
                  /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT      164 a    147 c    187 g    157 t
ORIGIN

  Query Match             14.2%;  Score 71;  DB 29;  Length 655;
  Best Local Similarity   70.3%;  Pred. No. 0.056;
  Matches  109;  Conservative    0;  Mismatches   45;  Indels    1;  Gaps    1;

Qy        287 GAGTTGGCCTTCTGACTTGCTTTGA-CAATGGAATGTAGTACAAATGACACTGTGCAACT 345

Db        462 GAGCTGGCCCTGTGACTTGCTTTGACCAATTGAATGGAGTAGAAGTGGTGCTCTGTGGTT 521

Qy        346 TTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGAT 405

Db        522 TCCAAGATTACGTCTTAAGAGGCCTTGCAGATTCCATTTTTGCCCTCTTGGGAGCCAGCT 581

Qy        406 GCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGA 440

Db        582 ACTATGTAAATAAGTCTGGACTATCCTGCTGGAAA 616


RESULT 6
AQ351210
LOCUS       AQ351210                 608 bp    DNA     linear   GSS 07-MAY-1999
DEFINITION  RPCI11-120H19.TV RPCI-11 Homo sapiens genomic clone RPCI-11-120H19,
            genomic survey sequence.
ACCESSION   AQ351210
VERSION     AQ351210.1  GI:4178545
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 608)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: T7
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..608
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7545882"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-120H19"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      176 a    146 c    126 g    160 t
ORIGIN

  Query Match             14.1%;  Score 70.6;  DB 28;  Length 608;
  Best Local Similarity   57.8%;  Pred. No. 0.065;
  Matches  189;  Conservative    0;  Mismatches  124;  Indels   14;  Gaps    3;

Qy        161 TGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCC- 219

Db        133 TGTGGTGGATTATTTGCAAAGACGGCATCAACAATTCCTCCCATCCCTATAGGTATGCAT 192

Qy        220 CCTTTGCAATGTGACTTTGCTACTTCTCTATCAA-----------GATGTGGAGCTTATT 268

Db        193 CCTATGGAATGAGACCTTGTAGCTCCTCTTATCAAGCTCCTCCTTGTAGCTCCTCTTATT 252

Qy        269 TTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACA 328

Db        253 TTTCCACCCCTTAAATTTTGGCTGGCCTGTGAGTTACTTGGACCCATAGAATGCAACAGA 312

Qy        329 AATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACAC 388

Db        313 AATGGCATTGCATGACATCCGAGCCGTGGCATCAAGAGGACTTACAGATTCTTCTCTCAC 372

Qy        389 TCTCTTGGAAA--CCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATG 446

Db        373 CCTCTTAGACAGTGATGCTACCATGTGGGGAAGCTCAAGCTAGCCTGCTGGAGGACAAGT 432

Qy        447 TCCCAGCTAATAGCCACAATCAACCTC 473

Db        433 TCACATGCAACAGAGATGAGCCAGCCC 459


RESULT 7
AQ614358
LOCUS       AQ614358                 555 bp    DNA     linear   GSS 15-JUN-1999
DEFINITION  HS_5126_B1_G10_T7A RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate-702 Col=19 Row=N, genomic survey sequence.
ACCESSION   AQ614358
VERSION     AQ614358.1  GI:5075634
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 555)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 702  row: N  column: 19
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 555.
FEATURES             Location/Qualifiers
     source          1..555
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="Plate=702 Col=19 Row=N"
                     /sex="male"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     Male blood DNA was isolated from one randomly chosen donor
```

```
                and partially digested with a combination of EcoRI and
                EcoRI Methylase. Size selected DNA was cloned into the
                pBACe3.6 vector at EcoRI sites"
BASE COUNT      134 a    133 c    126 g    154 t      8 others
ORIGIN

  Query Match             13.5%;  Score 67.8;  DB 28;  Length 555;
  Best Local Similarity   66.5%;  Pred. No. 0.18;
  Matches  115;  Conservative    0;  Mismatches   52;  Indels    6;  Gaps    1;

Qy        142 AAATTCTACACATAATCATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCC 201
Db        100 AACTTTTCCTCATATCTGCTATGGGAGATTGCTTCCAGTGTTGGCCACAACAATGCCT-- 157

Qy        202 TATCCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGA 261
Db        158 ----CTCACCCACATGCCCTAGGCAGTGTGACTTTGCCCCCTTTCAATCAATAAGTAAA 213

Qy        262 GCTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAA 314
Db        214 GCTTATTCCCCCTCCTCTTGAATCTGAGCTAGGCTTCTGACTAGCTTTCACCA 266


RESULT 8
BX103709
LOCUS       BX103709                 722 bp    mRNA    linear   EST 06-FEB-2003
DEFINITION  BX103709 Soares retina N2b4HR Homo sapiens cDNA clone
            IMAGp998A05437 ; IMAGE:220300, mRNA sequence.
ACCESSION   BX103709
VERSION     BX103709.1  GI:27845840
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 722)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998A05437.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1..722
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998A05437 ; IMAGE:220300"
                     /sex="male"
                     /tissue_type="retina"
                     /dev_stage="55 year old"
                     /lab_host="DH10B (ampicillin resistant)"
                     /clone_lib="Soares retina N2b4HR"
                     /note="Organ: eye; Vector: pT7T3D (Pharmacia) with a
                     modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
                     strand cDNA was primed with a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGCTTTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was size selected, ligated to Eco RI
                     adapters (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of a modified pT7T3 vector
                     (Pharmacia). The retinas were obtained from a 55 year old
                     Caucasian  and total cellular poly(A)+ RNA was extracted 6
                     hrs after their removal. The retina RNA  was kindly
                     provided by Roderick R. McInnes M.D. Ph.D. from the
                     University  of Toronto. Library constructed by Bento
                     Soares and M.Fatima Bonaldo. "
BASE COUNT      193 a    182 c    178 g    169 t
ORIGIN

  Query Match             13.5%;  Score 67.4;  DB 13;  Length 722;
  Best Local Similarity   61.2%;  Pred. No. 0.21;
  Matches  183;  Conservative    0;  Mismatches  101;  Indels   15;  Gaps    4;

Qy        217 GCCCCTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCAT 275
Db        338 GCTGCTTACTAATGTGACTTTGCCACTGTTCTGGGCAAGAGGTGGAGTCTGTTTCCCTAC 397

Qy        276 ATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACA 335
Db        398 CCCTTGAACCTGGGTTCGCC-TGGGACTACTTTGGCCAATAGAATGTGGCAGAAGTGATG 456

Qy        336 CTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTG 395
Db        457 TTATGCAAATTCTGACATTAGGCCTTAAGAGACCTTGCAGCTTTCACGTTTACTGTCCTG 516

Qy        396 GAAACCAG-------ATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGA-----CA 442
Db        517 GAACACAGTCCTGAGACTGCTATGTAAGGAAGCCAGCCTACTGGAGGATGAAAGAACCAA 576

Qy        443 TATGTCCCAGCTAATAGCCACAATCAACCTCTGAACATATGAATGAGGCTAGCTAGGCC 501
Db        577 TGTGTCCCCGCCGACAGCCAAGACCAACTGCCAGACATGTAAATGAGCCATCTTGGACC 635


RESULT 9
AL703889/c
LOCUS       AL703889                 588 bp    mRNA    linear   EST 22-MAR-2002
DEFINITION  DKFZp686B1128_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
            DKFZp686B1128 5', mRNA sequence.
ACCESSION   AL703889
VERSION     AL703889.1  GI:19687244
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 588)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann
            ,S.)
  JOURNAL   Unpublished
COMMENT     Contact: Ottenwaelder B
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No sl sequence
            available.
            This clone (DKFZp686B1128) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..588
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686B1128"
                     /tissue_type="human skeletal muscle"
                     /dev_stage="adult"
```

```
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
BASE COUNT      174 a    105 c    147 g    162 t
ORIGIN

  Query Match            13.3%;  Score 66.4;  DB 9;  Length 586;
  Best Local Similarity   58.7%;  Pred. No. 0.3;
  Matches  169;  Conservative    0;  Mismatches  116;  Indels    3;  Gaps    3;

Qy        161 TGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCC 220

Db        447 TGAAGTAGATTGTTTGTATAGAAGGCCACAA-ATTTCCTCCTACCTTTCTATTCATGTCC 389

Qy        221 CTTTGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGAGCTTATTTTCCCATATAT 279

Db        388 TTTTGCTTCCTGACTTTGCTGCTTCTCCCATCAAGAACTGGAGTCCATTTCCACCTCCTC 329

Qy        280 TGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGT 339

Db        328 TTGAATCTGCCCTGGCCTTGTCTTGCTTTGACCAGTCAAATGAAATGGAAGTAATTTCAC 269

Qy        340 GCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAA 399

Db        268 ACAACTACTAAGTCTAGGTCTCCAGGTGCCTTGCAATTTTCACTCTCACTCTCCTACCTG 209

Qy        400 CCAGATGCAATGTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATGT 447

Db        208 AGATCT-TTTTGTAAATAAGTCTGAGACAGGCTGCTATAGAGATTAGT 162


RESULT 10
AI041146/c
LOCUS       AI041146                 477 bp    mRNA    linear   EST 27-AUG-1998
DEFINITION  ov69f02.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1642587
            3', mRNA sequence.
ACCESSION   AI041146
VERSION     AI041146.1  GI:3280340
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 477)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
            , Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 974   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1..477
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1642587"
                     /sex="male"
                     /lab_host="DH10B"
                     /clone_lib="Soares_testis_NHT"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was prepared from mRNA obtained from Clontech Laboratories
                     , Inc., and primed with  a Not I - oligo(dT) primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
                     Double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization to Cot5, and was
                     constructed  by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      126 a     98 c    107 g    146 t
ORIGIN

  Query Match            13.1%;  Score 65.6;  DB 9;  Length 477;
  Best Local Similarity   63.5%;  Pred. No. 0.4;
  Matches  132;  Conservative    0;  Mismatches   74;  Indels    2;  Gaps    2;

Qy        234 CTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGG 293

Db        410 CTTTGCAACTCTCCCATTATGAGGTGGAGCCTATTTATTCATTCCTTTGACCTGGACTGA 351

Qy        294 CCTTCTGACTTGCTTTGAC-AATGGAATGTAGTACAAATGACACTGTGCAACTT-TGGAT 351

Db        350 CTTTGTGACTTGCTTTGAATAAAGGAATAGAACAGAAGTGATGTTGCATAAGTTCTGGAG 291

Qy        352 TTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAATG 411

Db        290 TTTAGGCCTCAAGAGAGCTTACAGTTTCCACCATCGCTCTCTTGGAACATTGTCTACCAT 231

Qy        412 TAAAGAAGTCAGGGCTATCCTGCTAGAG 439

Db        230 CTAAGGAAGTCAGTCTATCCTACTGGAG 203


RESULT 11
BF929327/c
LOCUS       BF929327                 474 bp    mRNA    linear   EST 19-JAN-2001
DEFINITION  IL2-NT0202 081200-298-C08 NT0202 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF929327
VERSION     BF929327.1  GI:12327455
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 474)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL2&t2=IL2-NT0202-
            081200-298-C08&t3=2000-12-08&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 59
            High quality sequence stop: 443.
FEATURES             Location/Qualifiers
     source          1..474
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
```

```
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="NT0202"
                     /note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      117 a    102 c    107 g    147 t      1 others
ORIGIN

  Query Match             13.0%;  Score 65;  DB 10;  Length 474;
  Best Local Similarity   57.3%;  Pred. No. 0.5;
  Matches  138;  Conservative    0;  Mismatches  100;  Indels    3;  Gaps    1;

Qy        234 CTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGG 293
Db        324 CTCTGCAGCTCTTCTATCAAGAGGTGCAATCTCTTTCCTGAAACATTGAATCTGGCTCAG 265

Qy        294 CCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTT 353
Db        264 CGTGTGCCTTGATTTGACCAGCAGAACAGAGCAGAAGTGATACTATGCAATTGCCCAGCC 205

Qy        354 TAGGTTTCGAGAGAACTTACAC---CTTCCACTCACACTCTCTTGGAAACCAGATGCAAT 410
Db        204 TAAGCTTCAAAAGAGTCCTCGTAGGCTTCCGTTCCCGCACACGGATCTCTCCACCTCTGT 145

Qy        411 GTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATGTCCCAGCTAATAGCCACAATCAAC 470
Db        144 GTGAAGAAGTCCAGGCTGGCTTGCTGGAGACATGTGGCCCAGCTAACAGCCAGCACCAAC 85

Qy        471 C 471
Db         84 C 84


RESULT 12
AQ080797
LOCUS       AQ080797                 401 bp    DNA     linear   GSS 20-AUG-1998
DEFINITION  CIT-HSP-2367H8.TF CIT-HSP Homo sapiens genomic clone 2367H8,
            genomic survey sequence.
ACCESSION   AQ080797
VERSION     AQ080797.1  GI:3441981
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 401)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K.,
            Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and
            Venter,J.C.
  TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: CIT-HSP-2367H8.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
            Seq primer: M13-21
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..401
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="2367H8"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
BASE COUNT      101 a     83 c     98 g    119 t
ORIGIN

  Query Match             12.9%;  Score 64.8;  DB 28;  Length 401;
  Best Local Similarity   63.0%;  Pred. No. 0.53;
  Matches  116;  Conservative    0;  Mismatches   67;  Indels    1;  Gaps    1;

Qy        170 TTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAAT 229
Db         75 TCATTTGCAAAGATGGTCTCAATAATTTCTCCCATGTCTATGGGCAGGTCTTTTTGCAGT 134

Qy        230 GTGACTTTGCTACT-TCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGA 288
Db        135 GTGACTTTGCCACTCTTTCTGTCAAGAAGTAGAGTTTATTTTCCATCCTCTTGCATCTAG 194

Qy        289 GTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTG 348
Db        195 TCTGGTCTTGTACTTGTTTTATCAAGTAAAATGTGGTAGAAGTGACTAGGGGATTCCTGA 254

Qy        349 GATT 352
Db        255 GCTT 258


RESULT 13
AK046951
LOCUS       AK046951                1335 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
            enriched library, clone:B930007I11 product:unknown EST, full insert
            sequence.
ACCESSION   AK046951
VERSION     AK046951.1  GI:26091826
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
```

```
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5
  AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1335)
  AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..1335
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:B930007I11"
                     /db_xref="taxon:10090"
                     /clone="B930007I11"
                     /tissue_type="cerebellum"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days neonate"
     misc_feature    1..1335
                     /note="unknown EST (GB|BI080508, evidence: BLASTN, 99%,
                     match=465)"
BASE COUNT      396 a    302 c    293 g    344 t

ORIGIN

  Query Match             12.9%;  Score 64.8;  DB 11;  Length 1335;
  Best Local Similarity   57.1%;  Pred. No. 0.55;
  Matches  137;  Conservative    0;  Mismatches  102;  Indels    1;  Gaps    1;

Qy        159 ATTGTGCTAAATTACTTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGC 218
                                                                          
Db        497 ACTGTAGTACATTGTTTGCAGAGATGGCCACAATAACTCCTCCCATTCCTGTATGTACAC 556

Qy        219 CCCTTTGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATA 278
                                                                          
Db        557 CCCTTGCAATGTGACTTTACCTTCTTCCCATGAAGGGGTGGCATATCTTTCCTCACTTT 616

Qy        279 TTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTG 338
                                                                          
Db        617 TTGAATCATTGCCATTTTTGTGGTTTGCTTAGACCACTGTTAGTTGGCCAAA-GCGACGT 675

Qy        339 TGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAA 398
                                                                          
Db        676 ATGTATACTGGAGCCTAGACTTCAAGAGGACATGCAGCTTTTGCCCTCACCCTCTAAGGA 735


RESULT 14
CC142611/c
LOCUS       CC142611                 865 bp    DNA     linear   GSS 16-APR-2003
DEFINITION  NDL.37D22.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
            NDL.37D22, genomic survey sequence.
ACCESSION   CC142611
VERSION     CC142611.1  GI:30011666
KEYWORDS    GSS.
SOURCE      Aedes aegypti (yellow fever mosquito)
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
REFERENCE   1  (bases 1 to 865)
  AUTHORS   Loftus,B., Shetty,J., Knudson,D. and Severson,D.
  TITLE     BAC end sequencing of Aedes aegypti
  JOURNAL   Unpublished
COMMENT     Other_GSSs: NDL.37D22.T7
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..865
                     /organism="Aedes aegypti"
                     /mol_type="genomic DNA"
                     /strain="Liverpool"
                     /db_xref="taxon:7159"
                     /clone="NDL.37D22"
                     /clone_lib="Notre Dame Liverpool"
                     /note="Vector: pECBAC1; Site_1: Hind III; The library was
                     prepared from whole body tissue of newly hatched L1 larvae
                     by David Severson at the University of Notre Dame and
                     Hongbin Zhang"
BASE COUNT      234 a    187 c    192 g    252 t
ORIGIN

  Query Match             12.7%;  Score 63.8;  DB 29;  Length 865;
  Best Local Similarity   65.1%;  Pred. No. 0.78;
  Matches  110;  Conservative    0;  Mismatches   57;  Indels    2;  Gaps    1;

Qy        205 CCTCATATATATGCCCCTTTGCAATGTGACTTTGCTACTTC--TCTATCAAGATGTGGAG 262
                                                                          
Db        824 CCCTGTATGCACATCTCTTTGCAATATCACTTTTAAAGTTCCTACCATCAAGAAGTGCAG 765
```

```
Qy        263 CTTATTTTCCCATATATTGCACTAGAGTTGGCCTTCTGACTTGCTTTGACAATGGAATGT 322
              |||| | |    | | |    |   |||||| ||  ||| | | |  ||    || ||
Db        764 CCTATTTCTCTAATCTTGGAATCTGGACTGGCCTTGTGACTTGCTTGGCCAACAAAATGT 705

Qy        323 AGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTT 371
              || |||| ||  || |   ||  |     ||| |  ||| | | |
Db        704 ATTAGAAATGACAGTGGGCCAGTTCTGAGCCTAGGCCTCAAGAGGAGTT 656


RESULT 15
BX116065/c
LOCUS       BX116065                 497 bp    mRNA    linear   EST 07-FEB-2003
DEFINITION  BX116065 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
            IMAGp998E035152 ; IMAGE:2080090, mRNA sequence.
ACCESSION   BX116065
VERSION     BX116065.1  GI:27880373
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 497)
  AUTHORS   Ebert,L., Heil,C., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
  TITLE     Human UnigeneSet - RZPD3
  JOURNAL   Unpublished
COMMENT     Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998E035152.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection:
            Human UnigeneSet - RZPD3 (RZPDLIB No.972)
            http://www.rzpd.de/CloneCards/cgi-
            bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Heubnerweg 6, D-14059 Berlin, Germany
            Tel: +49 30 32639 101
            Fax: +49 30 32639 111
            www.rzpd.de
            This clone is available royalty-free from RZPD;
            contact RZPD (clone@rzpd.de) for further information. Seq primer:
            M13r, Primer sequence: TTTCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1. .497
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998E035152 ; IMAGE:2080090"
                     /lab_host="DH10B"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      154 a     98 c    120 g    125 t
ORIGIN

  Query Match           12.7%;  Score 63.6;  DB 13;  Length 497;
  Best Local Similarity  56.4%;  Pred. No. 0.83;
  Matches  159;  Conservative    0;  Mismatches  119;  Indels    4;  Gaps    2;

Qy        174 TTGCAAAGATGGCCACAACAATTCCTCCTATCCTCATATATATGCCCCTTTGCAATGTGA 233
              ||   |||.| ||:. | ||| |||: |||| | |. :     |  .||| ||  |||
Db        468 TTTATAAGATAGCCAAATCAAGTCCTTCTATCCTTATGTGCCCACTCCTTTACAATGTGA 409

Qy        234 CTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTGCACTAGAGTTGG 293
              .|    |  :|| | : ||| : || : -  ||:| | | |     |    | |
Db        408 CTGCCTTTTTTCCC-ACCAAGAGGTAGGGTCTATTTTGCTACCCCCTCAGTCTGAACTAG 350

Qy        294 CCTTCTGACTTGCTTTGACAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTT 353
              ||       :  ||   |:|  | |  || |  |  |'   |||| ::.|
Db        349 CCCATGACTTGCTTTGACCAATCCAGTAGAGCAGAAGAGATGCTGTGTGACTTCTAGGCC 290

Qy        354 TAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCAGATGCAAT--- 410
              |    ||: |||   || || |||:  ||| :  |||:|   ||  .   |||
Db        289 TCAACTTCAAGAGGTGTTGCAGCTTCCGCTCTCGTTCTCTTTCAACACTTCTGCCATCAC 230

Qy        411 GTAAAGAAGTCAGGGCTATCCTGCTAGAGACATATGTCCCAG 452
              | ||||        ||   ||     ||  :  | | ' '|
Db        229 GTAAAGAAACTCAGTCTGGTCTTCCTTAGCCATTTATTCAAG 188


Search completed: October 24, 2003, 22:51:01
Job time : 1050.9 secs
```

GenCore version 5.1.6

DM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1384.09 Seconds
(without alignments)
14808.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500
Perfect score: 501
Sequence: 1 ggttttgcgtgagtttctta.....atcagagaggtggattggaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2886711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgc_hum:*
40: em_htgc_mus:*
41: em_htgc_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		501	100.0	110000	6	AX093471	AX093471 Sequence
2		501	100.0	110000	9	HSA246003	AC246003 Homo sapi
3		501	100.0	162692	9	CNS01DS8	AL121658 BAC seque
4		501	100.0	185281	2	AC011232	AC011232 Homo sapi
c	5	445.6	88.9	173106	9	AC068985	AC068985 Homo sapi
6		445.6	88.9	349938	6	AX647882	AX647882 Sequence
7		445	88.8	196963	9	AC009623	AC009623 Homo sapi
c	8	442.4	88.3	132868	2	AC021323	AC021323 Homo sapi
c	9	442.4	88.3	155323	9	AC022598	AC022598 Homo sapi
c	10	442.4	88.3	158404	9	AC090094	AC090094 Homo sapi
11		442.4	88.3	161644	9	AC025133	AC025133 Homo sapi
12		442.4	88.3	169505	9	CNS01DTV	AL132987 Human chr
13		441.8	88.2	99966	9	AC080012	AC080012 Homo sapi
c	14	440.8	88.0	91875	9	AL512286	AL512286 Human DNA
15		440.8	88.0	93089	9	AY308744	AY308744 Homo sapi
16		440.8	88.0	110000	2	AL831785_0	AL831785 Homo sapi
c	17	440	87.8	187525	2	AL354678	AL354678 Homo sapi
18		439.2	87.7	64965	9	AC006962	AC006962 Homo sapi
19		439.2	87.7	98056	9	AC006061	AC006061 Homo sapi
c	20	439.2	87.7	136631	9	AC109808	AC109808 Homo sapi
c	21	439.2	87.7	152832	9	AC069509	AC069509 Homo sapi
c	22	439.2	87.7	152951	9	AC026371	AC026371 Homo sapi
c	23	439.2	87.7	165590	9	AL512885	AL512885 Human DNA
24		439.2	87.7	215819	9	AC140490	AC140490 Homo sapi
c	25	439.2	87.7	222880	6	AX647883	AX647883 Sequence
c	26	439.2	87.7	243887	2	AC024898	AC024898 Homo sapi
c	27	438.8	87.6	157348	2	AP002446	AP002446 Homo sapi
28		438.8	87.6	166960	2	AC015958	AC015958 Homo sapi
29		438.8	87.6	176224	2	AP000898	AP000898 Homo sapi
30		438.8	87.6	184440	9	AP000919	AP000919 Homo sapi
c	31	438.8	87.6	185178	2	AP000937	AP000937 Homo sapi
c	32	438.4	87.5	155837	9	AL445495	AL445495 Human DNA
33		438.4	87.5	349993	6	AX647271	AX647271 Sequence
c	34	438.2	87.5	86141	9	BX323845	BX323845 Human DNA
35		438.2	87.5	182627	2	AC024262	AC024262 Homo sapi
36		438.2	87.5	182776	9	AC091078	AC091078 Homo sapi
37		438.2	87.5	191605	9	AC027670	AC027670 Homo sapi
c	38	437.6	87.3	41770	9	AC127090	AC127090 Homo sapi
c	39	437.6	87.3	73175	9	AC111193	AC111193 Homo sapi
c	40	437.6	87.3	74085	9	AC110780	AC110780 Homo sapi
41		437.6	87.3	79884	9	AC109440	AC109440 Homo sapi
42		437.6	87.3	86588	9	AC010453	AC010453 Homo sapi
43		437.6	87.3	114420	9	AL158076	AL158076 Human DNA
44		437.6	87.3	139550	9	AC110062	AC110062 Homo sapi
45		437.6	87.3	145824	9	AC090667	AC090667 Homo sapi

ALIGNMENTS

RESULT 1
AX093471
LOCUS AX093471 110000 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 1 from Patent WO0118198.
ACCESSION AX093471
VERSION AX093471.1 GI:13509911
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Weissenbach,J. and Hazan,J.
TITLE Cloning, expression and characterisation of the spg4 gene
responsible for the most frequent form of autosomal spastic

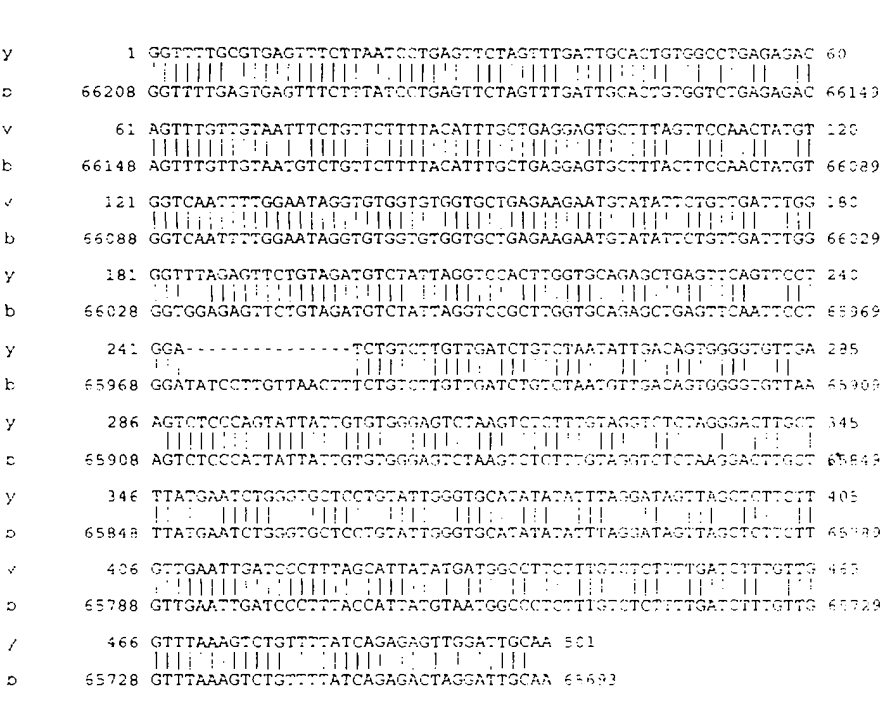
```
           paraplegia
  JOURNAL   Patent: WO 0118198-A 1 15-MAR-2001;
            CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES             Location/Qualifiers
     source          1. .110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     exon            9932. .10471
     intron          10472. .33718
     exon            33719. .33805
     intron          33806. .35748
     exon            35749. .35832
     intron          35833. .45022
     exon            45023. .45118
     intron          45119. .60863
     exon            60864. .61051
     intron          61052. .61927
     exon            61928. .62061
     intron          62062. .62344
     exon            62345. .62438
     intron          62439. .73173
     exon            73174. .73248
     intron          73249. .74633
     exon            74634. .74705
     intron          74706. .82788
     exon            82789. .82864
     intron          82865. .83102
     exon            83103. .83194
     intron          83195. .83334
     exon            83335. .83414
     intron          83415. .88129
     exon            88130. .88172
     intron          88173. .89561
     exon            89562. .89641
     intron          89642. .91162
     exon            91163. .91233
     intron          91234. .93443
     exon            93444. .93484
     intron          93485. .100599
     exon            100600. .102009
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match           100.0%;  Score 501;  DB 6;  Length 110000;
  Best Local Similarity  100.0%;  Pred. No. 2.8e-118;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50000 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 50059

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50060 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 50119

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50120 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 50179

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50180 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 50239

Qy        241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50240 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 50299

Qy        301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50300 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 50359

Qy        361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50360 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 50419

Qy        421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50420 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 50479

Qy        481 TATCAGAGAGTTGGATTGCAA 501
              |||||||||||||||||||||
Db      50480 TATCAGAGAGTTGGATTGCAA 50500


RESULT 2
HSA246003
LOCUS       HSA246003             110000 bp    DNA     linear   PRI 08-MAR 2000
DEFINITION  Homo sapiens Spast gene for spastin protein.
ACCESSION   AJ246003
VERSION     AJ246003.1  GI:6273432
KEYWORDS    Spast gene; spastin protein; SPG4-linked hereditary spastic
            paraplegia.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a new AAA protein, is altered in the most frequent form of
            autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. 23 (3), 296-303 (1999)
  MEDLINE   20055425
   PUBMED   10610178
REFERENCE   2
  AUTHORS   Fonknechten,N., Mavel,D., Byrne,P., Davoine,C., Cruaud,C.,
            Boentsch,D., Samson,D., Coutinho,P., Hutchinson,M., McMonagle,P.,
            Burgunder,J., Tartaglione,A., Heinzlef,O., Feki,I., Deufel,T.,
            Parfrey,N., Brice,A., Fontaine,B., Prud'homme,J., Weissenbach,J.,
            Durr,A. and Hazan,J.
  TITLE     Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
  JOURNAL   Hum. Mol. Genet. 9 (4), 637-644 (2000)
  MEDLINE   20164302
   PUBMED   10699187
REFERENCE   3  (bases 1 to 110000)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
            Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT     E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The
            sequence is the result of the assembly of 2 BAC clones: R-336p14
            and
            563N4, respectively from RPCI-11 and CITB_978_SKB library.
FEATURES             Location/Qualifiers
     source          1. .110000
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2p21"
                     /clone="R-336P14"
                     /clone_lib="RPCI-11"
                     /note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
                     , BAC 563N4 contains Spast gene: exon 4 to exon 17"
     gene            9932. .102009
                     /gene="Spast"
     mRNA            9932. .102009
                     /gene="Spast"
                     /product="spastin protein"
     exon            9932. .10471
                     /gene="Spast"
                     /number=1
```

```
5'UTR           9932..10056
                /gene="Spast"
CDS             join(10057..10471,33719..33805,35749..35832,45023..45118,
                60864..61051,61928..62061,62345..62438,73174..73248,
                74634..74705,82789..82864,83103..83194,83335..83414,
                88130..88172,89562..89641,91163..91233,93444..93484,
                100600..100722)
                /gene="Spast"
                /codon_start=1
                /product="spastin protein"
                /protein_id="CAB60208.1"
                /db_xref="GI:6273493"
                /translation="MNSPGGRGKKKGSGGASNPVPPRPPPPCLAPAPPAAGPAPPPES
                PHKRNLYYFSYPLFVGFALLRLVAFHLGLLFVWLCQRFSRALMAAKRSSGAAPAPASA
                SAPAPVPGGEAERVRVFHKQAFEYISIALRIDEDEKAGQKEQAVEWYKKGIEELEKGI
                AVIVTGQGEQCERARRLQAKMMTNLVMAKDRLQLLEKMQPVLPFSKSQTDVYNDSTNL
                ACRNGHLQSESGAVPKRKDPLTHTSNSLPRSKTVMKTGSAGLSGHHRAPSYSGLSMVS
                GVKQGSGPAPTTHKGTPKTNRTNKPSTPTTATRKKKDLKNFRNVDSNLANLIMNEIVD
                NGTAVKFDDIAGQDLAKQALQEIVILPSLRPELFTGLRAPARGLLLFGPPGNGKTMLA
                KAVAAESNATFFNISAASLTSKYVGEGEKLVRALFAVARELQPSIIFIDEVDSLLCER
                REGEHDASRRLKTEFLIEFDGVQSAGDDRVLVMGATNRPQELDEAVLRRFIKRVYVSL
                PNEETRLLLLKNLLCKQGSPLTQKELAQLARMTDGYSGSDLTALAKDAALGPIRELKP
                EQVKNMSASEMRNIRLSDFTESLKKIKRSVSPQTLEAYIRWNKDFGDTTV"
intron          10472..33718
                /gene="Spast"
                /number=1
exon            33719..33805
                /gene="Spast"
                /number=2
intron          33806..35748
                /gene="Spast"
                /number=2
exon            35749..35832
                /gene="Spast"
                /number=3
intron          35833..45022
                /gene="Spast"
                /number=3
exon            45023..45118
                /gene="Spast"
                /number=4
intron          45119..60863
                /gene="Spast"
                /number=4
exon            60864..61051
                /gene="Spast"
                /number=5
intron          61052..61927
                /gene="Spast"
                /number=5
exon            61928..62061
                /gene="Spast"
                /number=6
intron          62062..62344
                /gene="Spast"
                /number=6
exon            62345..62438
                /gene="Spast"
                /number=7
intron          62439..73173
                /gene="Spast"
                /number=7
exon            73174..73248
                /gene="Spast"
                /number=8
intron          73249..74633
                /gene="Spast"
                /number=8
exon            74634..74705
                /gene="Spast"
                /number=9
intron          74706..82788
                /gene="Spast"
                /number=9
exon            82789..82864
                /gene="Spast"
                /number=10
intron          82865..83102
                /gene="Spast"
                /number=10
exon            83103..83194
                /gene="Spast"
                /number=11
intron          83195..83334
                /gene="Spast"
                /number=11
exon            83335..83414
                /gene="Spast"
                /number=12
intron          83415..88129
                /gene="Spast"
                /number=12
exon            88130..88172
                /gene="Spast"
                /number=13
intron          88173..89561
                /gene="Spast"
                /number=13
exon            89562..89641
                /gene="Spast"
                /number=14
intron          89642..91162
                /gene="Spast"
                /number=14
exon            91163..91233
                /gene="Spast"
                /number=15
intron          91234..93443
                /gene="Spast"
                /number=15
exon            93444..93484
                /gene="Spast"
                /number=16
intron          93485..100599
                /gene="Spast"
                /number=16
exon            100600..102009
                /gene="Spast"
                /number=17
3'UTR           100721..102009
                /gene="Spast"
polyA_signal    101973..101978
                /gene="Spast"
BASE COUNT    30622 a  21640 c  22817 g  34921 t
ORIGIN

  Query Match             100.0%;  Score 501;  DB 9;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-118;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50000 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 50059

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50060 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 50119

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50120 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 50179

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50180 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 50239
```

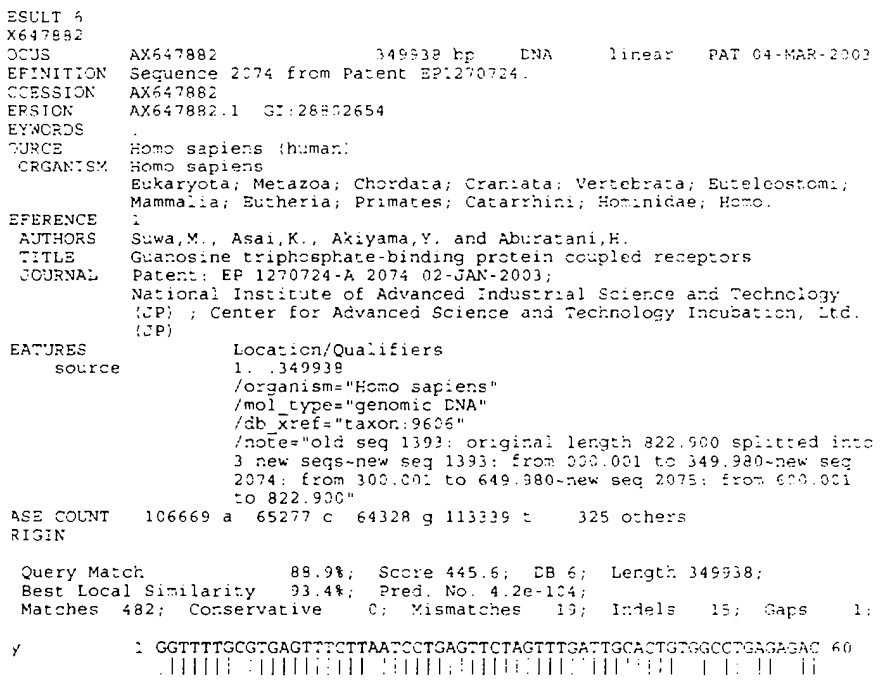
```
Qy       241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
Db     50240 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 50299

Qy       301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
Db     50300 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 50359

Qy       361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420
Db     50360 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 50419

Qy       421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480
Db     50420 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 50479

Qy       481 TATCAGAGAGTTGGATTGCAA 501
Db     50480 TATCAGAGAGTTGGATTGCAA 50500


RESULT 3
CNS01DS8
LOCUS       CNS01DS8               162692 bp    DNA     linear   PRI 16-APR-2002
DEFINITION  BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04
            of library RPCI-11 from chromosome 2 of Homo sapiens (Human).
ACCESSION   AL121658
VERSION     AL121658.4  GI:20218783
KEYWORDS    SPG4 genomic DNA interval.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 162692)
  AUTHORS   Hazan,J., Fonknechten,N., Mavel,D., Paternotte,C., Samson,D.,
            Artiguenave,F., Davoine,C.S., Cruaud,C., Durr,A., Wincker,P.,
            Brottier,P., Cattolico,L., Barbe,V., Burgunder,J.M.,
            Prud'Homme,J.F., Brice,A., Fontaine,B., Heilig,R. and
            Weissenbach,J.
  TITLE     Spastin, a novel AAA protein, is altered in the most frequent form
            of autosomal dominant spastic paraplegia
  JOURNAL   Nat. Genet. (1999) In press
REFERENCE   2  (bases 1 to 162692)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     On Apr 19, 2002 this sequence version replaced gi:20160242.
FEATURES             Location/Qualifiers
     source          1..162692
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="563N04"
                     /clone_lib="RPCI-11"
BASE COUNT    44981 a  32017 c  32887 g  52792 t     15 others
ORIGIN

  Query Match             100.0%;  Score 501;  DB 9;  Length 162692;
  Best Local Similarity   100.0%;  Pred. No. 2.7e-116;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db     31259 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 31318

Qy        61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db     31319 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 31378

Qy       121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db     31379 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 31438

Qy       181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db     31439 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 31498

Qy       241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
Db     31499 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 31558

Qy       301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
Db     31559 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 31618

Qy       361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420
Db     31619 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 31678

Qy       421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480
Db     31679 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 31738

Qy       481 TATCAGAGAGTTGGATTGCAA 501
Db     31739 TATCAGAGAGTTGGATTGCAA 31759


RESULT 4
AC011232
LOCUS       AC011232               185281 bp    DNA     linear   HTG 10-MAR-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
            7 unordered pieces.
ACCESSION   AC011232
VERSION     AC011232.7  GI:13270720
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185281)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-OCT-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 10, 2001 this sequence version replaced gi:9799811.

            ---------------- Genome Center ----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ------------- Project Information -------------
            Center project name: H_NH0078E13
            -------------- Summary Statistics -------------
            Sequencing vector: M13; 57%
            Sequencing vector: plasmid; 40%
            Chemistry: Dye-primer ET; 48% of reads
            Chemistry: Dye-terminator Big Dye; 50% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 182798 bases at least Q40
            Consensus quality: 183538 bases at least Q30
            Consensus quality: 184045 bases at least Q20
            Insert size: 168000; agarose-fp
            Insert size: 184681; sum-of-contigs
            Quality coverage: 7.07 in Q20 bases; agarose-fp
            Quality coverage: 7.33 in Q20 bases; sum-of-contigs
            -----------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
```

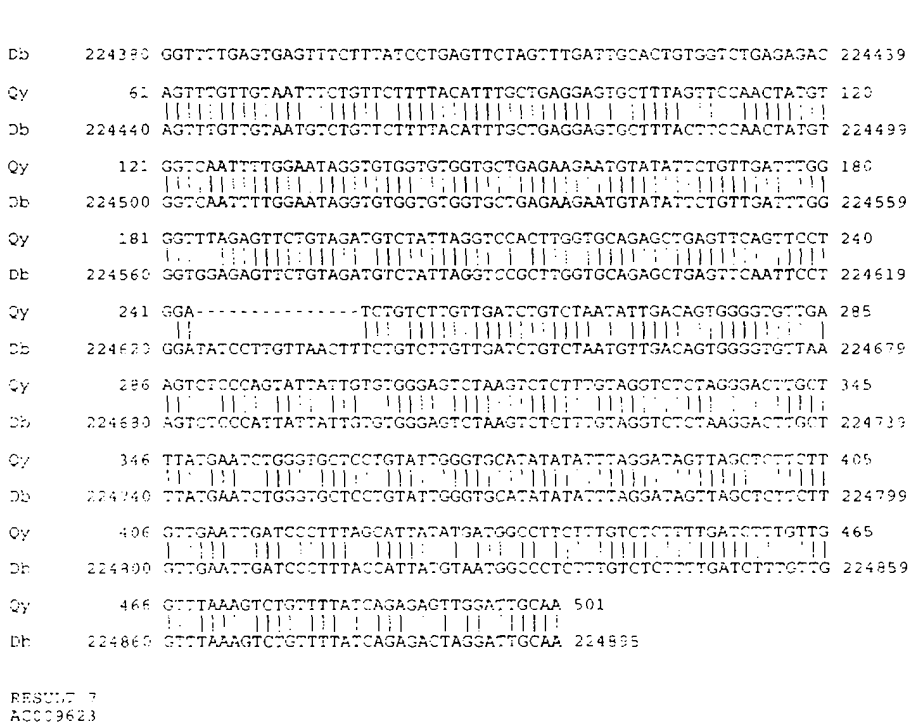
```
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1985: contig of 1985 bp in length
            *     1986     2085: gap of unknown length
            *     2086     8183: contig of 6098 bp in length
            *     8184     8283: gap of unknown length
            *     8284    22741: contig of 14458 bp in length
            *    22742    22841: gap of unknown length
            *    22842    41446: contig of 18605 bp in length
            *    41447    41546: gap of unknown length
            *    41547    69182: contig of 27636 bp in length
            *    69183    69282: gap of unknown length
            *    69283   112131: contig of 42849 bp in length
            *   112132   112231: gap of unknown length
            *   112232   185281: contig of 73050 bp in length.
FEATURES             Location/Qualifiers
     source          1. .185281
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-78B13"
     misc_feature    1. .1985
                     /note="assembly_name:Contig11"
     misc_feature    2086. .8183
                     /note="assembly_name:Contig12"
     misc_feature    8284. .22741
                     /note="assembly_name:Contig13"
     misc_feature    22842. .41446
                     /note="assembly_name:Contig14"
     misc_feature    41547. .69182
                     /note="assembly_name:Contig15"
     misc_feature    69283. .112131
                     /note="assembly_name:Contig16"
     misc_feature    112232. .185281
                     /note="assembly_name:Contig17"
BASE COUNT    52719 a  36957 c  37187 g  57816 t    602 others
ORIGIN

  Query Match           100.0%;  Score 501;  DB 2;  Length 185281;
  Best Local Similarity  100.0%;  Pred. No. 2.7e-118;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81395 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 81454

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81455 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 81514

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81515 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 81574

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81575 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 81634

Qy        241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81635 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 81694

Qy        301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81695 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 81754

Qy        361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81755 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 81814

Qy        421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81815 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 81874

Qy        481 TATCAGAGAGTTGGATTGCAA 501
              |||||||||||||||||||||
Db      81875 TATCAGAGAGTTGGATTGCAA 81895


RESULT 5
AC068985/c
LOCUS       AC068985              173106 bp    DNA     linear   PRI 01-DEC-2001
DEFINITION  Homo sapiens 3 BAC RP11-217E22 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC068985
VERSION     AC068985.17  GI:16554308
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 173106)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.F., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,C.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.C., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,C., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            He,X., Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M.,
            Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S.,
            Hume,J., Jackson,L.R., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,Z., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Ogun,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubckan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 173106)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
```

```
EFERENCE   3  (bases 1 to 173106)
 AUTHORS   Worley,K.C.
 TITLE     Direct Submission
 JOURNAL   Submitted (31-OCT-2001) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
EFERENCE   4  (bases 1 to 173106)
 AUTHORS   Worley,K.C.
 TITLE     Direct Submission
 JOURNAL   Submitted (01-DEC-2001) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
OMMENT     On Oct 31, 2001 this sequence version replaced gi:16418054.
           INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
           gc-help@bcm.tmc.edu

           CLONE LENGTH: This sequence does not necessarily represent the
           entire insert of this clone. Overlapping regions of clones are only
           sequenced and submitted once, so the sequence for the remainder of
           the insert may be found in the record for the adjacent clones.
           Overlapping clones are noted at the beginning and end of the
           Features listing.

           ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550) searches
           of a local database that includes entries from dbSTS, GDB, and
           local mapping efforts.
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
           unpublished.) for Human and Mouse sequences.
              Genes and Region of sequence similarity are identified by BLAST
           (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
           EST and cDNA sequences. Genes demonstrate at least two exons
           flanked by consensus splice sites that maintained sequence
           continuity across the splice junctions. Sequences that are not
           identical matches are annotated as similar.

           SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
           standard of double strand coverage with a minimum of 2 clones and 2
           reads with no ambiguities or 2 chemistries with a minimum of 2
           clones and 3 reads with no ambiguities. If the sequence quality for
           a region does not meet this standard, it will be indicated in the
           annotation as Low Coverage.

           QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
           standards - estimated error rate less than 1 per 10,000 bases.
           Reports of lowest quality individual bases and measures of base
           quality are listed below. Description of the metrics can be found
           at URL:
           http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

           QUALSTAT-REPORT.
EATURES             Location/Qualifiers
     source         1..173106
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="3"
                    /clone="RP11-217E22"
     misc_feature   1..2001
                    /note="overlaps bases 129142..131142 of clone AC073156"
                    /function="clone overlap"
     repeat_region  99..507
                    /rpt_family="MSTC"
     repeat_region  2220..2342
                    /rpt_family="MIR"
     repeat_region  2951..3613
                    /rpt_family="L1M5b"
     STS            4275..4612
                    /standard_name="24462"
     repeat_region  4553..4589
                    /rpt_family="(TG)n"
     repeat_region  complement(4816..5031)
                    /rpt_family="MIR"
     repeat_region  5240..5296
                    /rpt_family="(TA)n"
     repeat_region  5353..5391
                    /rpt_family="(TAGA)n"
     repeat_region  5527..5920
                    /rpt_family="MER93"
     repeat_region  6402..6521
                    /rpt_family="MIR"
     repeat_region  6827..7020
                    /rpt_family="MIR"
     repeat_region  8733..8757
                    /rpt_family="AT_rich"
     repeat_region  complement(8842..9151)
                    /rpt_family="AluJo"
     repeat_region  9165..9187
                    /rpt_family="AT_rich"
     repeat_region  9767..9931
                    /rpt_family="MSTD"
     repeat_region  10272..10426
                    /rpt_family="MIR"
     repeat_region  12479..12828
                    /rpt_family="AT_rich"
     repeat_region  12730..13185
                    /rpt_family="MER89"
     repeat_region  14101..14151
                    /rpt_family="AT_rich"
     repeat_region  14576..14626
                    /rpt_family="AT_rich"
     repeat_region  14758..14951
                    /rpt_family="HAL1"
     repeat_region  15200..15558
                    /rpt_family="HAL1"
     repeat_region  complement(15909..16144)
                    /rpt_family="MER8"
     repeat_region  16260..16295
                    /rpt_family="A-rich"
     repeat_region  16365..16395
                    /rpt_family="AT_rich"
     repeat_region  complement(17961..18026)
                    /rpt_family="L2"
     repeat_region  complement(18027..18332)
                    /rpt_family="AluY"
     repeat_region  complement(18333..18346)
                    /rpt_family="L2"
     repeat_region  complement(19180..19424)
                    /rpt_family="L2"
     repeat_region  complement(19425..19914)
                    /rpt_family="L1PA5"
     repeat_region  complement(19815..20200)
                    /rpt_family="L2"
     repeat_region  complement(20532..20938)
                    /rpt_family="MLT2E"
     repeat_region  20959..21521
                    /rpt_family="HERV16"
     repeat_region  21645..22064
                    /rpt_family="HERV16"
     repeat_region  22356..22486
                    /rpt_family="LTR50"
     repeat_region  22509..22554
                    /rpt_family="CT-rich"
     repeat_region  complement(22555..22857)
                    /rpt_family="AluSx"
     repeat_region  22856..22873
                    /rpt_family="CT-rich"
     repeat_region  22965..23767
                    /rpt_family="HERVL"
     repeat_region  24216..24609
                    /rpt_family="LTR16A"

Query Match             88.9%;  Score 445.6;  DB 9;  Length 173106;
Best Local Similarity   93.4%;  Pred. No. 4.3e-104;
Matches  482;  Conservative    0;  Mismatches   19;  Indels   15;  Gaps    1;
```

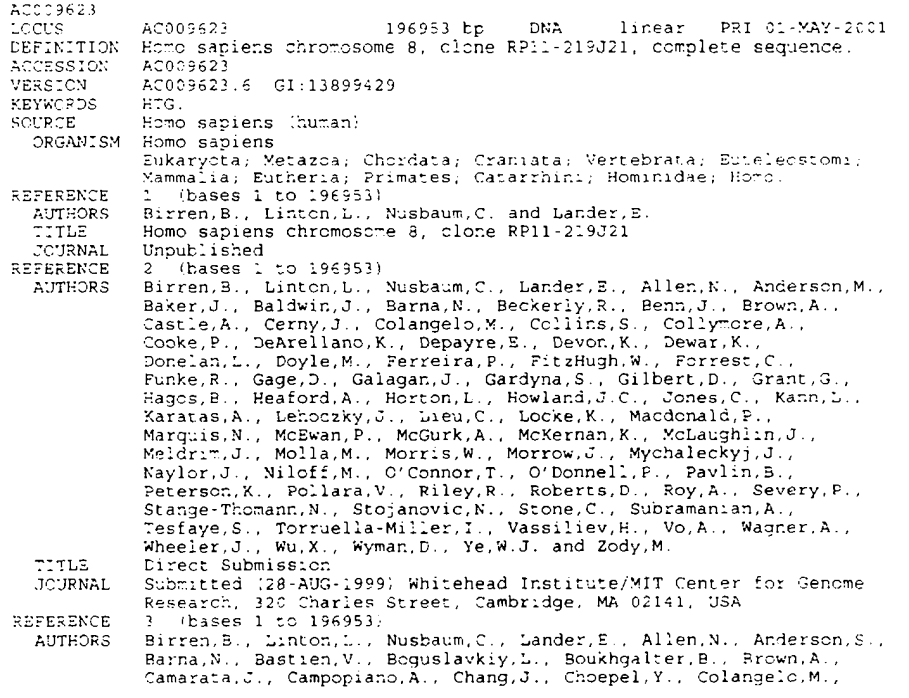


```
ESULT 6
X647882
OCUS       AX647882               349938 bp    DNA     linear   PAT 04-MAR-2003
EFINITION  Sequence 2074 from Patent EP1270724.
CCESSION   AX647882
ERSION     AX647882.1  GI:28802654
EYWORDS    .
OURCE      Homo sapiens (human)
 RGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE   1
 AUTHORS   Suwa,M., Asai,K., Akiyama,Y. and Aburatani,H.
 TITLE     Guanosine triphosphate-binding protein coupled receptors
 JOURNAL   Patent: EP 1270724-A 2074 02-JAN-2003;
           National Institute of Advanced Industrial Science and Technology
           (JP) ; Center for Advanced Science and Technology Incubation, Ltd.
           (JP)
EATURES             Location/Qualifiers
     source         1. .349938
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /note="old seq 1393: original length 822.900 splitted into
                    3 new seqs-new seq 1393: from 000.001 to 349.980-new seq
                    2074: from 300.001 to 649.980-new seq 2075: from 600.001
                    to 822.900"
ASE COUNT   106669 a  65277 c  64328 g 113339 t    325 others
RIGIN

 Query Match           89.9%;  Score 445.6;  DB 6;  Length 349938;
 Best Local Similarity   93.4%;  Pred. No. 4.2e-104;
 Matches  482;  Conservative    0;  Mismatches   19;  Indels   15;  Gaps    1;
```





```
RESULT 7
AC009623
LOCUS       AC009623               196953 bp    DNA     linear   PRI 01-MAY-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-219J21, complete sequence.
ACCESSION   AC009623
VERSION     AC009623.6  GI:13899429
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 196953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-219J21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 196953)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Chang,J., Choepel,Y., Colangelo,M.,
```



```
            Collins,S., Collymore,A., Cooke,P., DeArellano,K., Dewar,K.,
            Diaz,C.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L., Hulme,W.,
            Iliev,I., Johnson,R., Jones,C., Karatas,A., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lin,S.,
            MacLean,C., Macdonald,P., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
            Norman,C.H., O'Conner,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Sougnez,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,C., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 1, 2001 this sequence version replaced gi:12313793.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2082
                Center clone name: 219_J_21
            --------------.

FEATURES             Location/Qualifiers
     source          1..196953
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-219J21"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1031..1514)
                     /rpt_family="MER66B"
     repeat_region   1950..1897
                     /rpt_family="(TTATG)n"
     repeat_region   complement(2080..2207)
                     /rpt_family="L1ME"
     repeat_region   2281..2311
                     /rpt_family="AT_rich"
     repeat_region   complement(2508..2630)
                     /rpt_family="FLAM_C"
     repeat_region   complement(2791..3085)
                     /rpt_family="AluSx"
     repeat_region   complement(3288..4068)
                     /rpt_family="L1MB8"
     repeat_region   complement(4069..4366)
                     /rpt_family="AluSc"
     repeat_region   complement(4367..4824)
                     /rpt_family="L1MB8"
     repeat_region   complement(5100..5217)
                     /rpt_family="AluSg/x"
     repeat_region   5233..6096
                     /rpt_family="L2"
     repeat_region   6151..6455
                     /rpt_family="MLT1I"
     repeat_region   6583..6694
                     /rpt_family="L2"
     repeat_region   complement(6698..6840)
                     /rpt_family="L1MEd"
     repeat_region   complement(6876..7080)
                     /rpt_family="L1MEd"
     repeat_region   complement(7081..7377)
                     /rpt_family="AluY"
     repeat_region   complement(7378..7459)
                     /rpt_family="L1MEd"
     repeat_region   complement(7581..7786)
                     /rpt_family="L1MEd"
     repeat_region   complement(8027..8143)
                     /rpt_family="L1"
     repeat_region   8434..8458
                     /rpt_family="AT_rich"
     repeat_region   complement(8581..9162)
                     /rpt_family="L1MB3"
     repeat_region   9277..9447
                     /rpt_family="MER5B"
     repeat_region   complement(9498..9650)
                     /rpt_family="L1"
     repeat_region   10365..10426
                     /rpt_family="CT-rich"
     repeat_region   10614..10642
                     /rpt_family="AT_rich"
     repeat_region   10643..10953
                     /rpt_family="AluSg1"
     unsure          complement(10924..10931)
                     /note="Less than 30 qual SNGL region"
     repeat_region   complement(11143..11438)
                     /rpt_family="AluJb"
     repeat_region   11510..11614
                     /rpt_family="MIR"
     repeat_region   11640..12135
                     /rpt_family="MLT2B2"
     repeat_region   complement(12480..12919)
                     /rpt_family="MSTB"
     repeat_region   13237..13353
                     /rpt_family="L2"
     repeat_region   complement(15268..15576)
                     /rpt_family="AluSp"
     repeat_region   complement(15742..15855)
                     /rpt_family="L1MA4A"
     repeat_region   complement(15856..16163)
                     /rpt_family="AluY"
     repeat_region   complement(16164..16923)
                     /rpt_family="L1MA4A"
     repeat_region   complement(16924..17209)
                     /rpt_family="AluSx"
     repeat_region   complement(17210..19516)
                     /rpt_family="L1MA4A"
     repeat_region   complement(19517..19829)
                     /rpt_family="AluSg"
     repeat_region   complement(19830..20438)
                     /rpt_family="L1MA4A"
     repeat_region   complement(20439..20534)
                     /rpt_family="MER3C"
     repeat_region   20535..20838
                     /rpt_family="AluSg"
     repeat_region   complement(20839..20959)
                     /rpt_family="MER3C"
     repeat_region   complement(20960..23361)
                     /rpt_family="L1MA4A"
     repeat_region   complement(23384..23566)
                     /rpt_family="L1MA4A"
     repeat_region   23694..23865
                     /rpt_family="L2"
     repeat_region   23924..23969
                     /rpt_family="AT_rich"
     repeat_region   complement(23976..24087)
                     /rpt_family="MER2"
     repeat_region   complement(24093..30094)
                     /rpt_family="L1PA3"
     unsure          24211..24223
                     /note="Less than 30 qual SNGL region"
     repeat_region   complement(30095..30229)
                     /rpt_family="AluJb"
     repeat_region   complement(30231..30324)
```

```
                    /rpt_family="MER2"
     repeat_region  30325..30443
                    /rpt_family="L2"
     repeat_region  complement(30635..30800)
                    /rpt_family="AluJb"
     repeat_region  30837..30924
                    /rpt_family="MIR"
     repeat_region  complement(31020..31105)
                    /rpt_family="MIR"
     repeat_region  31416..31455
                    /rpt_family="(TG)n"
     repeat_region  complement(31456..31774)
                    /rpt_family="AluY"
     repeat_region  32488..32696
                    /rpt_family="MIR"
     repeat_region  34213..34394
                    /rpt_family="MIR3"
     repeat_region  complement(34399..34684)

  Query Match             88.8%;  Score 445;  DB 9;  Length 196953;
  Best Local Similarity   93.0%;  Pred. No. 6.1e-104;
  Matches  466;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      27397 GGTTTTGAGTGAGTTGCTTAATCCTGAGTTCCAGTTTGATTGCACTGTGGTCTGAGAGAC 27456

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      27457 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 27516

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      27517 GGTCAATTTTGGAATAGATGTGGTGTGGTGCTGAAAAAAATGTGTATTCTGTTGATTTGG 27576

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      27577 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTCCT 27636

Qy        241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
Db      27637 GGGTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAAAGTCTCCCATTATTA 27696

Qy        301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
Db      27697 TTGTGTGGGAGTCTAAGTCTCTTTGTAAGTCACTCAGGACTTGCTTTATGAATCTGGGTG 27756

Qy        361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420
Db      27757 CTCCTGTATTGGGTGCCTATACATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 27816

Qy        421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480
Db      27817 TTATCATTATGCAGTGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTATTT 27876

Qy        481 TATCAGAGAGTTGGATTGCAA 501
Db      27877 TATCAGAGACTAAGATTGCAA 27897


RESULT 8
AC021323/c
LOCUS       AC021323               132868 bp    DNA     linear   HTG 13-MAY-2001
DEFINITION  Homo sapiens clone RP11-23M4, WORKING DRAFT SEQUENCE, 2 ordered
            pieces.
ACCESSION   AC021323
VERSION     AC021323.5  GI:14030007
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 132868)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-23M4
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 132868)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Ferestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 13, 2001 this sequence version replaced gi:12229503.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4398
                Center clone name: 23_M_4
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 3% of reads
                Sequencing vector: Plasmid; n/a; 97% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 132571 bases at least Q40
                Consensus quality: 132698 bases at least Q30
                Consensus quality: 132753 bases at least Q20
                Insert size: 127000; agarose-fp
                Insert size: 132768; sum-of-contigs
                Quality coverage: 20.3 in Q20 bases; agarose-fp
                Quality coverage: 19.5 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1   130383: contig of 130383 bp in length
            *   130384   130483: gap of 100 bp
            *   130484   132868: contig of 2385 bp in length.
FEATURES             Location/Qualifiers
     source          1..132868
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-23M4"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..130383
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    130484..132868
                     /note="assembly_fragment
```

```
                clone_end:T7
                vector_side:right"
BASE COUNT    41079 a  26651 c  25911 g  39127 t    100 others
ORIGIN

  Query Match            88.3%;  Score 442.4;  DB 2;  Length 132868;
  Best Local Similarity  93.0%;  Pred. No. 2.9e-103;
  Matches  480;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      45102 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 45043

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      45042 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTGTACTTCCAACTATGT 44983

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      44982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 44923

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      44922 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTTCAGAGCTGAGTTCAATTCCT 44863

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      44862 GGATATCCTTGTTAACTTTCTGTCTTGATGATCTGTCTAATGTTGACAGTGGGGTGTTGA 44803

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      44802 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 44743

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      44742 TTATGAATCTGGGTGCTCCTGTATTGGGTACATATATATTTAGGATAGTTAGTTCTTCTT 44683

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      44682 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTATGTCTCTTTTGATCTTTGTTG 44623

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      44622 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 44587


RESULT 9
AC022598/c
LOCUS       AC022598               155323 bp    DNA     linear   PRI 11-DEC-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-119K7, complete sequence.
ACCESSION   AC022598
VERSION     AC022598.10  GI:17488692
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 155323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-119K7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 155323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenester,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 155323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 155323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   5  (bases 1 to 155323)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
```

```
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
OMMENT      On Dec 11, 2001 this sequence version replaced gi:15421013.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L4197
                Center clone name: 119_K_7
            --------------.
EATURES              Location/Qualifiers
     source          1..155323
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-119K7"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(274..575)
                     /rpt_family="AluSq"
     repeat_region   690..713
                     /rpt_family="(T)n"
     repeat_region   2402..2703
                     /rpt_family="AluJb"
     repeat_region   2740..3060
                     /rpt_family="AluJo"
     repeat_region   complement(3497..3616)
                     /rpt_family="L1ME"
     repeat_region   3775..3796
                     /rpt_family="AT_rich"
     repeat_region   complement(3917..4219)
                     /rpt_family="AluJo"
     repeat_region   complement(4292..4488)
                     /rpt_family="L1MC4"
     repeat_region   complement(4518..4625)
                     /rpt_family="L1MC4"
     repeat_region   complement(4626..4858)
                     /rpt_family="L1MC4a"
     repeat_region   complement(4861..5234)
                     /rpt_family="L1MC4a"
     repeat_region   5235..5271
                     /rpt_family="(CA)n"
     repeat_region   5360..5486
                     /rpt_family="FLAM_A"
     repeat_region   complement(5882..6168)
                     /rpt_family="AluSq"
     repeat_region   complement(6282..6613)
                     /rpt_family="AluJo"
     repeat_region   complement(6787..7063)
                     /rpt_family="AluSx"
     repeat_region   complement(8034..8334)
                     /rpt_family="AluSx"
     repeat_region   9606..9637
                     /rpt_family="AT_rich"
     repeat_region   10117..10175
                     /rpt_family="L2"
     repeat_region   10268..10422
                     /rpt_family="MLT1J2"
     repeat_region   complement(11063..11135)
                     /rpt_family="MER119"
     repeat_region   11138..11265
                     /rpt_family="(TA)n"
     repeat_region   complement(11301..11428)
                     /rpt_family="MER119"
     repeat_region   complement(11566..11669)
                     /rpt_family="FLAM_C"
     repeat_region   12321..12474
                     /rpt_family="MER5B"
     repeat_region   12568..12785
                     /rpt_family="MLT1K"
     repeat_region   complement(12895..13070)
                     /rpt_family="AluJo"
     repeat_region   complement(13071..13379)
                     /rpt_family="AluSc"
     repeat_region   complement(13380..13520)
                     /rpt_family="AluJo"
     repeat_region   13546..14035
                     /rpt_family="MLT1C"
     repeat_region   complement(14038..14251)
                     /rpt_family="L3"
     repeat_region   14361..14379
                     /rpt_family="MER5A"
     repeat_region   complement(14380..14676)
                     /rpt_family="AluJo"

  Query Match             88.3%;  Score 442.4;  DB 9;  Length 155323;
  Best Local Similarity   93.0%;  Pred. No. 2.9e-103;
  Matches  480;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db      58112 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 58053

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db      58052 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 57993

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db      57992 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 57933

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db      57932 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 57873

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db      57872 GGATATCCTTGTAAATTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 57813

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db      57812 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 57753

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db      57752 TTATGAATCTGGCTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 57693

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db      57692 CTTGAATTGATCCCTTCACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 57633

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db      57632 GTTTAAAGTCTGTTTTATCAGAGACCAGGGTTGCAA 57597

RESULT 10
AC090094/c
```

```
LOCUS       AC090094              158404 bp    DNA     linear   PRI 15-MAR-2002
DEFINITION  Homo sapiens chromosome 8, clone RP11-252C19, complete sequence.
ACCESSION   AC090094
VERSION     AC090094.5  GI:18482342
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 158404)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-252C19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 158404)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 158404)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 15, 2002 this sequence version replaced gi:14626341.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L12758
                Center clone name: 252_C_19
            --------------.
FEATURES             Location/Qualifiers
     source          1. .158404
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-252C19"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   733. .864
                     /rpt_family="L2"
     repeat_region   complement(1037. .1417)
                     /rpt_family="MLT2D"
     repeat_region   complement(1697. .1825)
                     /rpt_family="L1ME"
     repeat_region   1826. .2120
                     /rpt_family="AluSx"
     unsure          1960. .1969
                     /note="<30 qual SNGL region"
     unsure          2000. .2020
                     /note="<30 qual SNGL region"
     unsure          2030. .2046
                     /note="<30 qual SNGL region"
     unsure          2086. .2096
                     /note="<30 qual SNGL region"
     repeat_region   complement(2121. .2899)
                     /rpt_family="L1ME"
     unsure          2153. .2157
                     /note="<30 qual SNGL region"
     repeat_region   2988. .3249
                     /rpt_family="MLT1F"
     repeat_region   complement(5672. .5745)
                     /rpt_family="MER5B"
     repeat_region   complement(7472. .7524)
                     /rpt_family="L2"
     repeat_region   8965. .9104
                     /rpt_family="MIR"
     repeat_region   10056. .10077
                     /rpt_family="AT_rich"
     repeat_region   10154. .10511
                     /rpt_family="L2"
     repeat_region   10770. .10866
                     /rpt_family="L2"
     repeat_region   complement(11277. .11584)
                     /rpt_family="AluSx"
     repeat_region   complement(11650. .11891)
                     /rpt_family="L1ME1"
     repeat_region   complement(11994. .12920)
                     /rpt_family="L1MEc"
     repeat_region   complement(12926. .13195)
                     /rpt_family="AluSg"
     repeat_region   complement(13196. .13833)
                     /rpt_family="L1MEc"
     repeat_region   complement(13820. .15009)
                     /rpt_family="L1"
     repeat_region   complement(15010. .15280)
                     /rpt_family="L1PA5"
     repeat_region   complement(15281. .17128)
                     /rpt_family="L1"
     repeat_region   complement(17129. .17260)
                     /rpt_family="FLAM_C"
     repeat_region   complement(17261. .18749)
                     /rpt_family="L1"
     repeat_region   complement(18759. .18838)
                     /rpt_family="HAL1"
     repeat_region   complement(18852. .18923)
                     /rpt_family="L1MEc"
     repeat_region   complement(19633. .19968)
                     /rpt_family="MER44D"
```

```
   repeat_region   complement(22330. .22430)
                   /rpt_family="MIR"
   repeat_region   22683. .22788
                   /rpt_family="MIR"
   repeat_region   complement(23594. .23742)
                   /rpt_family="L2"
   repeat_region   23905. .23944
                   /rpt_family="(CA)n"
   repeat_region   complement(25236. .25754)
                   /rpt_family="L1ME1"
   repeat_region   complement(25768. .25920)
                   /rpt_family="L2"
   repeat_region   complement(26462. .26590)
                   /rpt_family="MIR3"
   repeat_region   26950. .26971
                   /rpt_family="AT_rich"
   repeat_region   complement(28241. .28452)
                   /rpt_family="L2"
   repeat_region   complement(28496. .28685)
                   /rpt_family="MIR"
   repeat_region   complement(28908. .29104)
                   /rpt_family="L2"
   repeat_region   29687. .29901
                   /rpt_family="FLAM_C"
   repeat_region   30765. .30857
                   /rpt_family="MIR"
   repeat_region   complement(30935. .31205)
                   /rpt_family="AluSx"
   repeat_region   31219. .31244
                   /rpt_family="(TAA)n"
   repeat_region   32601. .32635
                   /rpt_family="AT_rich"
   repeat_region   33555. .34025
                   /rpt_family="MLT1D"
   repeat_region   34037. .34565
                   /rpt_family="THE1C-int"
   repeat_region   34566. .34585
                   /rpt_family="(T)n"
   repeat_region   36347. .36601
                   /rpt_family="THE1C-int"
   repeat_region   36602. .36975
                   /rpt_family="THE1C"
   repeat_region   37072. .37095
                   /rpt_family="(T)n"
   repeat_region   complement(37155. .37423)
                   /rpt_family="L1PA5"
   repeat_region   37564. .37623
                   /rpt_family="AT_rich"
   repeat_region   complement(38380. .38821)
                   /rpt_family="MER74B"
   repeat_region   complement(38874. .39178)
                   /rpt_family="L2"
   repeat_region   complement(39345. .39549)
                   /rpt_family="L2"

 Query Match             88.3%;  Score 442.4;  DB 9;  Length 158404;
 Best Local Similarity   93.0%;  Pred. No. 2.9e-103;
 Matches  480;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    1;

y          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
b     136382 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 136323

y         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
b     136322 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTGTACTTCCAACTATGT 136263

y        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
b     136262 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 136203

y        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
```

```
Db     136202 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTTCAGAGCTGAGTTCAATTCCT 136143

Qy        241 GGA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db     136142 GGATATCCTTGTTAACTTTCTGTCTTGATGATCTGTCTAATGTTGACAGTGGGGTGTTGA 136083

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db     136082 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 136023

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     136022 TTATGAATCTGGGTGCTCCTGTATTGGGTACATATATATTTAGGATAGTTAGTTCTTGTT 135963

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db     135962 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTATGTCTCTTTTGATCTTTGTTG 135903

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db     135902 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 135867
```

```
RESULT 11
AC025133
LOCUS       AC025133              161644 bp    DNA     linear   PRI 01-DEC-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-227F6, complete sequence.
ACCESSION   AC025133
VERSION     AC025133.6  GI:17149771
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 161644)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-227F6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 161644)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,C.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.C.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 161644)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
```

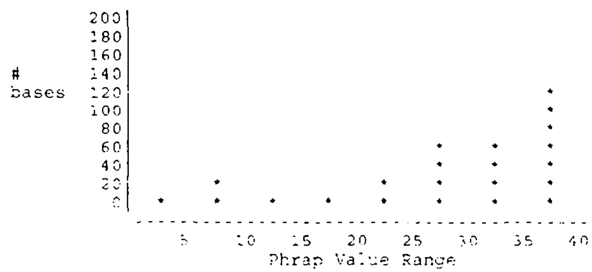
```
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lin,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPherson,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,S.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.C., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Nov 29, 2001 this sequence version replaced gi:16041527.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7841
                Center clone name: 227_F_6
            --------------.
FEATURES             Location/Qualifiers
     source          1. .161644
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-227F6"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   complement(1620. .1753)
                     /rpt_family="MIR"
     repeat_region   complement(1804. .1969)
                     /rpt_family="MIR3"
     repeat_region   complement(2309. .2500)
                     /rpt_family="MER5A"
     repeat_region   2922. .2987
                     /rpt_family="(TG)n"
     repeat_region   3640. .3812
                     /rpt_family="GA-rich"
     repeat_region   3829. .3875
                     /rpt_family="GA-rich"
     repeat_region   3917. .4078
                     /rpt_family="GA-rich"
     repeat_region   complement(4121. .4739)
                     /rpt_family="L1MA8"
     repeat_region   complement(7998. .8143)
                     /rpt_family="MIR"
     repeat_region   complement(8197. .8303)
                     /rpt_family="AluSp"
     repeat_region   9579. .9628
                     /rpt_family="(TG)n"
     repeat_region   10046. .10072
                     /rpt_family="(TAAA)n"
     repeat_region   10156. .10266
                     /rpt_family="MER5B"
     repeat_region   complement(10626. .10969)
                     /rpt_family="L2"
     repeat_region   complement(11083. .11311)
                     /rpt_family="L2"
     repeat_region   11340. .11524
                     /rpt_family="L1MA8"
     repeat_region   11584. .11640
                     /rpt_family="(TA)n"
     repeat_region   complement(11647. .11685)
                     /rpt_family="L1MA8"
     repeat_region   complement(11689. .11887)
                     /rpt_family="L1MA8"
     repeat_region   complement(11887. .12214)
                     /rpt_family="L1MA8"
     repeat_region   12215. .12443
                     /rpt_family="L1PA5"
     repeat_region   complement(12444. .13281)
                     /rpt_family="L1MA8"
     repeat_region   13291. .13311
                     /rpt_family="AT_rich"
     repeat_region   complement(13343. .13515)
                     /rpt_family="L2"
     repeat_region   14013. .14089
                     /rpt_family="L1MA8"
     repeat_region   15105. .16664
                     /rpt_family="L1MD2"
     repeat_region   complement(16665. .16980)
                     /rpt_family="MER2"
     repeat_region   16981. .18155
                     /rpt_family="L1MD2"
     repeat_region   18172. .18899
                     /rpt_family="L1MD2"
     repeat_region   19262. .19284
                     /rpt_family="AT_rich"
     repeat_region   complement(19371. .21157)
                     /rpt_family="L1PA3"
     repeat_region   21299. .21404
                     /rpt_family="L1MC4"
     repeat_region   complement(21416. .21813)
                     /rpt_family="AluSq"
     repeat_region   21869. .21908
                     /rpt_family="MIR3"
     repeat_region   21918. .22013
                     /rpt_family="MIR3"
     unsure          22015
                     /note="one of the reads contains an A"
     repeat_region   22562. .22589
                     /rpt_family="(CAAA)n"
     repeat_region   complement(23312. .23612)
                     /rpt_family="AluSx"
     repeat_region   24398. .24420
                     /rpt_family="(CAAT)n"
     repeat_region   complement(24805. .25228)
                     /rpt_family="MSTA"
     repeat_region   complement(25753. .25943)
                     /rpt_family="MER31B"
     repeat_region   complement(25972. .26137)
                     /rpt_family="MER31B"
     repeat_region   complement(26155. .26267)
                     /rpt_family="L2"
     repeat_region   complement(26495. .26619)
                     /rpt_family="MER20B"
     unsure          complement(28926. .28931)
                     /note="<30 qual SNGL region"
     unsure          29086. .29269
                     /note="single clone coverage"
     unsure          29131. .29141
                     /note="<30 qual single clone coverage"
     repeat_region   29203. .29498
                     /rpt_family="AluSq"
     repeat_region   29712. .29751
                     /rpt_family="(TG)n"
     repeat_region   complement(30588. .30705)
                     /rpt_family="L2"
     repeat_region   complement(31556. .31748)
                     /rpt_family="MIR"
     repeat_region   32920. .33038
                     /rpt_family="CT-rich"
     repeat_region   34145. .34172
                     /rpt_family="AT_rich"
     repeat_region   complement(34176. .34349)
                     /rpt_family="L1PB1"
```

```
    repeat_region   complement(34350..34644)
                    /rpt_family="AluY"
    repeat_region   complement(34645..35403)
                    /rpt_family="L1PB1"
    repeat_region   complement(36372..37043)
                    /rpt_family="L1MC3"

 Query Match             88.3%;  Score 442.4;  DB 9;  Length 161644;
 Best Local Similarity   93.0%;  Pred. No. 2.9e-103;
 Matches  480;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    1;

y          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
b     142079 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 142138

y         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
b     142139 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTGTACTTCCAACTATGT 142198

y        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
b     142199 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 142258

y        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
b     142259 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTTCAGAGCTGAGTTCAATTCCT 142318

y        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
b     142319 GGATATCCTTGTTAACTTTCTGTCTTGATGATCTGTCTAATGTTGACAGTGGGGTGTTGA 142378

y        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
b     142379 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 142438

y        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
b     142439 TTATGAATCTGGGTGCTCCTGTATTGGGTACATATATATTTAGGATAGTTAGTTCTTCTT 142498

y        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
b     142499 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTATGTCTCTTTTGATCTTTGTTG 142558

y        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
      142559 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 142594

ESULT 12
NS01DTV
OCUS       CNS01DTV              169505 bp    DNA     linear   PRI 01-JUN-2001
EFINITION  Human chromosome 14 DNA sequence BAC R-188I13 of library RPCI-11
           from chromosome 14 of Homo sapiens (Human), complete sequence.
CCESSION   AL132987
ERSION     AL132987.4  GI:14284836
EYWORDS    HTG; HTGS_ACTIVEFIN.
OURCE      Homo sapiens (human)
 ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE   1  (bases 1 to 169505)
 AUTHORS   Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
           Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
           Levy,M., Eckenberg,R., Bruls,T., deBerardinis,V., Cruaud,C.,
           Gyapay,G., Saurin,W. and Weissenbach,J.
 TITLE     Sequencing of the human chromosome 14
 JOURNAL   Unpublished
EFERENCE   2  (bases 1 to 169505)
 AUTHORS   Genoscope.
 TITLE     Direct Submission
 JOURNAL   Submitted (31-MAY-2001) Genoscope - Centre National de Sequencage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
COMMENT    On Jun 3, 2001 this sequence version replaced gi:8217887.
           --------------- Genome Center
           Center: Genoscope / Centre National de Sequencage
           Center code: GS
           Web site: http://www.genoscope.cns.fr/
           Contact: SeqRef@genoscope.cns.fr
           ---------------
           The following BAC sequence is oriented from the T7 to the SP6 end.
           Upstream BAC (overlapping the T7 end) : R-862G15 (AC=AL117192)
           Downstream BAC (overlapping the SP6 end) : R-807G16 --------------
           Summary Statistics
           Assembly program: Phrap; version 2.0
           Quality coverage: 8.35x in Q20 bases; sum-of-contigs
           ---------------
           ----------------------
           Overall quality chart :
           Range     :    bases
           0         :
           1  -  9   :
           10 - 19   :        3
           20 - 29   :        6
           30 - 39   :       43
           40 - 49   :      716
           50 - 59   :     1710
           60 - 69   :     5293
           70 - 79   :    19513
           80 - 89   :    59301
           90 - 99   :    82920
           ----------------------
           Percentage of bases with a quality value >= 40 : 99 %.
FEATURES             Location/Qualifiers
     source          1..169505
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="14"
                     /clone="R-188I13"
                     /clone_lib="RPCI-11"
     STS             45652..45764
                     /note="matching EMBL:F01540
                     RHdb:RH26869
                     dbSTS:STS6768
                     Identified using the e-PCR software (G. Schuler)"
     STS             47188..47291
                     /note="matching EMBL:G32424
                     RHdb:RH67113
                     dbSTS:STS47040
                     Identified using the e-PCR software (G. Schuler)"
BASE COUNT    47640 a   36003 c   37377 g   48485 t
ORIGIN

  Query Match             88.3%;  Score 442.4;  DB 9;  Length 169505;
  Best Local Similarity   93.0%;  Pred. No. 2.9e-103;
  Matches  480;  Conservative    0;  Mismatches   21;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      29793 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 29852

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      29853 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 29912

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      29913 GGTCAATTTTGGAATAAGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 29972

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      29973 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 30032

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
```

```
b       30033 GGATATCTGTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 30092

y         286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

b       30093 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 30152

y         346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

b       30153 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 30212

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Cb      30213 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCCCTTTGATCTTTGTTG 30272

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Cb      30273 GTTTAAAGTCTGTTTTATCAGCGACTAGGATTGCAA 30308
```

```
ESULT 13
C080012
OCUS        AC080012               99966 bp    DNA     linear   FRI 31-MAY-2002
EFINITION   Homo sapiens 12 BAC RP11-714K16 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
CCESSION    AC080012
ERSION      AC080012.20  GI:14290372
EYWORDS     HTG.
OURCE       Homo sapiens (human)
 RGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
EFERENCE    1  (bases 1 to 99966)
 AUTHORS    Muzny,D.M., Adams,C., Adio-Cduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
            Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.D., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
            Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
            Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
            Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
            Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
            Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Horsi,F.,
            Howard,S., Huber,J., Hulyk,S., Hume,C., Ioshikhes,I., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,S., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
            Martindale,A., Martinez,E., Massey,S., Mawhiney,S., McLeod,M.P.,
            Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
            Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
            Morris,S., Moser,M., Neal,D., Nelson,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
            Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
            Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
            Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
            Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
            Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
            Washington,C., Watlington,S., Williams,G., Williamson,A.,
            Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
            Zorrilla,S., Kucherlapati,R. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 99966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-SEP-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 99966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 99966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 99966)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 5, 2001 this sequence version replaced gi:14017474.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
               STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
               Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
               Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8086/quality.info/genbank.annotation.html.

            QUALSTAT-REPORT----------


            -- -- ------ --- Summary Statistics ------------------------------
            Contig length:                                  73346
            Phrap values in estimate:                       73208
            Average error rate (BCM-Phrap estimate):        8.64574e-05
            Fraction of Phrap values less than 40 :         0.00487652
            Number of consensus changing edits:             0
```

```
Number of N's in consensus :                        0

---------------- Consensus changing edits ------------------------
Position      Original+Context              Edited+Context

--------- Distribution of Quality < 40 Bases ------- ------------
```




```
---------------------------------------------------------------------
        Version: 1.01  qxfo.
EATURES          Location/Qualifiers
   source        1. .99966
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
                 /chromosome="12"
                 /clone="RP11-714K16"
   repeat_region complement(1. .94)
                 /rpt_family="MER104"
   repeat_region 394. .639
                 /rpt_family="L1MEc"
   repeat_region complement(672. .772)
                 /rpt_family="MIR"
   repeat_region 899. .1315
                 /rpt_family="L1M4"
   repeat_region 1322. .1536
                 /rpt_family="MER82"
   repeat_region 1537. .1817
                 /rpt_family="L1ME"
   repeat_region 1843. .2315
                 /rpt_family="L1ME"
   repeat_region 2480. .2637
                 /rpt_family="L2"
   repeat_region 2740. .3044
                 /rpt_family="AluSp"
   STS           2829. .2946
                 /standard_name="G02122"
                 /db_xref="dbSTS:7608"
   repeat_region 3853. .4076
                 /rpt_family="MLT1F"
   repeat_region 4803. .4830
                 /rpt_family="(A)n"
   repeat_region complement(5627. .5800)
                 /rpt_family="MER5A"
   repeat_region complement(6336. .6629)
                 /rpt_family="AluSp"
   repeat_region complement(6775. .6968)
                 /rpt_family="MIR"
   repeat_region 7444. .7475
                 /rpt_family="AT_rich"
   repeat_region 7654. .7761
                 /rpt_family="L1MB4"
   repeat_region 7857. .7878
                 /rpt_family="AT_rich"
   repeat_region 8597. .8637
                 /rpt_family="(TG)n"
   repeat_region 8728. .9784
                 /rpt_family="L1MB3"

  Query Match             88.2%;  Score 441.8;  DB 9;  Length 99966;
  Best Local Similarity   93.5%;  Pred. No. 4.1e-103;
  Matches  474;  Conservative    0;  Mismatches   27;  Indels    6;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      13495 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTCCACTGTGGTCTGAGAGAC 13554

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      13555 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAAGAGTGCTTTACTTCCAACTATGT 13614

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      13615 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAGAATGTATATTCTGTTGATTTGG 13674

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      13675 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGTCGAGTTCAATTCCT 13734

Qy        241 GG------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCA 294
Db      13735 GGTTAACTTTCAGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAAAGTCTCCCA 13794

Qy        295 GTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATC 354
Db      13795 TTATTATTGTGTGGGAGTCTAAGTCTCTTTCTAGGTCACTAAGGACTTGCTTTATGAATC 13854

Qy        355 TGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTG 414
Db      13855 TGGGTGCTCCTATATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTG 13914

Qy        415 ATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGT 474
Db      13915 ATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGT 13974

Qy        475 CTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      13975 CTGTTTTATCAGAGACTAGGATTGCAA 14001


RESULT 14
AL512286/c
LOCUS       AL512286               91875 bp    DNA     linear   PRI 02-FEB-2001
DEFINITION  Human DNA sequence from clone RP11-45J1 on chromosome X, complete
            sequence.
ACCESSION   AL512286
VERSION     AL512286.6  GI:12666312
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 91875)
  AUTHORS   Heath,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 5, 2001 this sequence version replaced gi:12578260.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated repeat sequence elements. Where the sequence is
```

```
            ambiguous, there is an annotation using the 'unsure' feature key.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome X, constructed by the Sanger Centre Chromosome X Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX
            RP11-45J1 is from the library RPCI-11.1 constructed by the group of
            Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-45J1 It may be shorter because we sequence overlapping
            sections only once, except for a 100 base overlap.
            The true left end of clone RP6-52J4 is at 91776 in this sequence.
            The true right end of clone RP3-525N14 is at 100 in this sequence.
FEATURES             Location/Qualifiers
     source          1..91875
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP11-45J1"
                     /clone_lib="RPCI-11.1"
     misc_feature    27383..27413
                     /note="Weak data"
BASE COUNT    26799 a  19653 c  18710 g  26713 t
ORIGIN

  Query Match             98.0%;  Score 440.8;  DB 9;  Length 91875;
  Best Local Similarity   92.8%;  Pred. No. 7.5e-103;
  Matches  479;  Conservative    0;  Mismatches   22;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      77478 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCATTGTGGTCTGAGAGAC 77419

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      77418 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 77359

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      77358 GGTCAATTTTGGAATAAGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 77299

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      77298 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTCCT 77239

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      77238 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTCATGTTGACAGTGGGGTGTTAA 77179

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      77178 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTGAGGACTTGCT 77119

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      77116 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 77059

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      77058 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 76999

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      76998 GTTTAAAGTCTGTTTTATCAAAGACTAGGATTGCAA 76963
```

```
RESULT 15
AY308744
LOCUS       AY308744               93089 bp    DNA     linear   PRI 05-JUN-2003
DEFINITION  Homo sapiens protein S (alpha) (PROS1) gene, partial cds.
ACCESSION   AY308744
VERSION     AY308744.1  GI:31415702
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 93089)
  AUTHORS   Rieder,M.J., Carrington,D.P., da Ponte,S.H., Hastings,N.C.,
            Ahearn,M.O., Kuldanek,S.A., Rajkumar,N., Toth,E.J., Yi,Q. and
            Nickerson,D.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-2003) Genome Sciences, University of Washington,
            1705 NE Pacific, Seattle, WA 98195, USA
COMMENT     To cite this work please use:  SeattleSNPs. NHLBI HL66682 Program
            for Genomic Applications, UW-FHCRC, Seattle, WA (URL:
            http://pga.gs.washington.edu).
FEATURES             Location/Qualifiers
     source          1..93089
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     variation       288
                     /frequency="0.22"
                     /replace="t"
     variation       430
                     /frequency="0.33"
                     /replace="g"
     variation       543..546
                     /frequency="0.01"
                     /replace=""
     variation       626
                     /frequency="0.01"
                     /replace="t"
     variation       804
                     /frequency="0.01"
                     /replace="a"
     variation       1194
                     /frequency="0.22"
                     /replace="t"
     variation       1368
                     /frequency="0.02"
                     /replace="a"
     gene            1906..>93089
                     /gene="PROS1"
     mRNA            join(1906..2130,48408..48565,51551..51575,65109..65195,
                     69673..69795,69901..70032,74886..75011,77248..77369,
                     79126..79241,82695..82884,89314..89481,90921..>93089)
                     /gene="PROS1"
                     /product="protein S (alpha)"
     CDS             join(2055..2130,48408..48565,51551..51575,65109..65195,
                     69673..69795,69901..70032,74886..75011,77248..77369,
                     79126..79241,82695..82884,89314..89481,90921..>91089)
                     /gene="PROS1"
                     /codon_start=1
                     /product="protein S (alpha)"
                     /protein_id="AAP45054.1"
                     /db_xref="GI:31415703"
                     /translation="MRVLGGRCGALLACLLLVLPVSEANFLSKQQASQVLVRKRRANS
                     LLEETKQGNLERECIEELCNKEEAREVFENDPETDYFYPKYLVCLRSFQTGLFTAARQ
                     STNAYPDLRSCVNAIPDQCSPLPCNEDGYMSCKDGKASFTCTCKPGWQGEKCEFDINE
                     CKDPSNINGGCSQICDNTPGSYHCSCKNGFVMLSNKKDCKDVDECSLKPSICGTAVCK
                     NIPGDFECECPEGYRYNLKSKSCEDIDECSENMCAQLCVNYPGGYTCYCDGKKGFKLA
                     QDQKSCEVVSVCLPLNLDTKYELLYLAEQFAGVVLYLKFRLPEISRFSAEFDFRTYDS
                     EGVILYAESIDHSAWLLIALRGGKIEVQLKNEHTSKITTGGDVINNGLWNMVSVEELE
                     HSISIKIAKEAVMDINKPGPLFKPENGLLETKVYFAGFPRKVESELIKPINPRLDGCI
                     RSWNLMKQGASGIKEIIQEKQNKHCLVTVEKGSYYPGSGIAQFHIDY"
     misc_feature    3462..6227
                     /gene="PROS1"
```

```
                /note="Region not scanned for variation"
variation       6952
                /gene="PROS1"
                /frequency="0.01"
                /replace="a"
variation       7052
                /gene="PROS1"
                /frequency="0.05"
                /replace=""
misc_feature    8396..12131
                /gene="PROS1"
                /note="Region not scanned for variation"
variation       12189
                /gene="PROS1"
                /frequency="0.01"
                /replace="c"
variation       12241
                /gene="PROS1"
                /frequency="0.04"
                /replace="g"
variation       12267
                /gene="PROS1"
                /frequency="0.02"
                /replace="t"
variation       12369
                /gene="PROS1"
                /frequency="0.01"
                /replace="a"
variation       12612
                /gene="PROS1"
                /frequency="0.01"
                /replace="g"
variation       13154
                /gene="PROS1"
                /frequency="0.14"
                /replace="a"
variation       13616
                /gene="PROS1"
                /frequency="0.07"
                /replace="t"
variation       14310
                /gene="PROS1"
                /frequency="0.03"
                /replace="a"
variation       14351
                /gene="PROS1"
                /frequency="0.01"
                /replace="a"
misc_feature    14486..17601
                /gene="PROS1"
                /note="Region not scanned for variation"
variation       17983
                /gene="PROS1"
                /frequency="0.02"
                /replace="g"
variation       18079
                /gene="PROS1"
                /frequency="0.04"
                /replace="a"
variation       18163
                /gene="PROS1"
                /frequency="0.01"
                /replace="c"
misc_feature    18349..22390
                /gene="PROS1"
                /note="Region not scanned for variation"
misc_feature    24564..26690
                /gene="PROS1"
                /note="Region not scanned for variation"
variation       26890
                /gene="PROS1"
                /frequency="0.07"
                /replace="c"
variation       27862
                /gene="PROS1"
                /frequency="0.03"
                /replace="a"
variation       28013
                /gene="PROS1"
                /frequency="0.01"
                /replace="t"
variation       28478
                /gene="PROS1"
                /frequency="0.01"
                /replace="a"
variation       28819
                /gene="PROS1"
                /frequency="0.06"
                /replace="c"
misc_feature    29123..32942
                /gene="PROS1"
                /note="Region not scanned for variation"
variation       33464
                /gene="PROS1"
                /frequency="0.01"
                /replace="t"
variation       33556
                /gene="PROS1"
                /frequency="0.01"
                /replace="c"
variation       33585
                /gene="PROS1"
                /frequency="0.33"
                /replace="a"
misc_feature    35058..39211
                /gene="PROS1"
                /note="Region not scanned for variation"
variation       40139
                /gene="PROS1"
                /frequency="0.03"
                /replace="g"
variation       40453
                /gene="PROS1"
                /frequency="0.01"
                /replace="a"
variation       40570
                /gene="PROS1"
                /frequency="0.01"
                /replace="t"
variation       40667
                /gene="PROS1"
                /frequency="0.15"
                /replace="t"
variation       40737
                /gene="PROS1"
                /frequency="0.01"
                /replace="c"
misc_feature    41422..45091
                /gene="PROS1"

Query Match             88.0%;  Score 440.8;  DB 9;  Length 93089;
Best Local Similarity   92.8%;  Pred. No. 7.5e-103;
Matches  479;  Conservative    0;  Mismatches   22;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db      23363 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 23422

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db      23423 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 23482

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db      23483 GGTCAATTTTGGAATAAGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 23542
```

```
Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db      23543 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCG 23602

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db      23603 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 23662

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db      23663 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 23722

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db      23723 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 23782

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db      23783 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 23842

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db      23843 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 23878
```

Search completed: October 24, 2003, 21:06:12
Job time : 1390.29 secs

GenCore version 5.1.6

DM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 126.251 Seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500
Perfect score: 501
Sequence: 1 ggttttgcgtgagtttctta.....atcagagagttggattgcaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	501	100.0	110000	22	AAF84800	Nucleotide sequenc
2	436	87.0	5065	21	AAC69111	Human secreted pro
3	436	87.0	5065	22	AAF33219	Human secreted pro
4	436	87.0	1691080	24	ABX08336	Human phosphodiest
5	434.4	86.7	9192	22	AAL38853	Human musculoskele
6	434.4	86.7	9192	23	ABK42359	Genomic sequence #
7	434.4	86.7	9192	25	ABX58841	cDNA encoding nove
c 8	432.8	86.4	1885	24	ABK92491	Human prostate spe
c 9	431.6	86.1	41599	21	AAZ35351	Cosmid including s
c 10	431.2	86.1	2277	23	AAS84982	DNA encoding novel
c 11	431.2	86.1	2606	23	AAS75832	DNA encoding novel
c 12	431.2	86.1	3316	23	AAS70259	DNA encoding novel
c 13	431.2	86.1	3316	23	AAS70804	DNA encoding novel
c 14	431.2	86.1	3316	23	AAS73840	DNA encoding novel
c 15	431.2	86.1	3316	23	AAS75000	DNA encoding novel
c 16	431.2	86.1	5405	23	AAS66524	DNA encoding novel
c 17	431.2	86.1	3588	23	AAS68192	DNA encoding novel
c 18	431.2	86.1	5130	23	AAS66528	DNA encoding novel
c 19	431.2	86.1	5130	23	AAS67239	DNA encoding novel
c 20	431.2	86.1	5130	23	AAS68166	DNA encoding novel
c 21	431.2	86.1	5130	23	AAS70242	DNA encoding novel
c 22	431.2	86.1	5130	23	AAS70789	DNA encoding novel
c 23	431.2	86.1	5130	23	AAS72822	DNA encoding novel
c 24	431.2	86.1	5130	23	AAS73825	DNA encoding novel
c 25	431.2	86.1	5130	23	AAS74984	DNA encoding novel
c 26	431.2	86.1	5130	23	AAS84060	DNA encoding novel
c 27	431.2	86.1	5130	23	AAS84966	DNA encoding novel
c 28	431.2	86.1	5130	23	AAS85098	DNA encoding novel
c 29	431.2	86.1	5130	23	AAS89455	DNA encoding novel
30	431.2	86.1	5131	23	AAS83326	DNA encoding novel
c 31	431.2	86.1	5210	23	AAS83526	DNA encoding novel
c 32	431.2	86.1	5213	23	AAS89597	DNA encoding novel
c 33	431.2	86.1	5231	23	AAS66699	DNA encoding novel
c 34	431.2	86.1	5450	23	AAS65193	DNA encoding novel
35	431.2	86.1	5676	22	AAL37391	Human musculoskele
36	431.2	86.1	5676	25	ABX60379	cDNA encoding nove
c 37	431.2	86.1	5773	23	AAS84049	DNA encoding novel
c 38	431.2	86.1	6194	23	AAS73339	DNA encoding novel
c 39	431.2	86.1	8065	23	AAS78738	DNA encoding novel
40	431.2	86.1	8739	22	AAK77090	Human immune/haema
c 41	431.2	86.1	9339	21	AAF20962	Human low adenosin
c 42	431.2	86.1	9339	21	AAA34840	Human adenosine re
c 43	431.2	86.1	10136	23	AAS75390	DNA encoding novel
c 44	431.2	86.1	10579	23	AAS69831	DNA encoding novel
c 45	431.2	86.1	22421	21	AAF20966	Human low adenosin

ALIGNMENTS

RESULT 1
AAF84800
ID AAF84800 standard; DNA; 110000 BP.
XX
AC AAF84800;
XX
DT 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 gene.
XX
KW Human; SPG4 gene; spastin; PSP-AD; gene therapy;
KW autosomal dominant familial spastic paraplegia; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 9932..102009
FT /*tag= a
FT /note= "contains introns"
FT exon 9932..10471
FT /*tag= b
FT /number= "1"
FT intron 10472..33718
FT /*tag= c
FT /number= "1"
FT exon 33719..33805
FT /*tag= d
FT /number= "2"
FT intron 33806..35748
FT /*tag= e
FT /number= "2"

```
FT   exon            35749..35832
FT                   /*tag=  f
FT                   /number= "3"
FT   intron          35833..45022
FT                   /*tag=  g
FT                   /number= "3"
FT   exon            45023..45118
FT                   /*tag=  h
FT                   /number= "4"
FT   intron          45119..60863
FT                   /*tag=  i
FT                   /number= "4"
FT   exon            60864..61051
FT                   /*tag=  j
FT                   /number= "5"
FT   intron          61052..61927
FT                   /*tag=  k
FT                   /number= "5"
FT   exon            61928..62061
FT                   /*tag=  l
FT                   /number= "6"
FT   intron          62062..62344
FT                   /*tag=  m
FT                   /number= "6"
FT   exon            62345..62438
FT                   /*tag=  n
FT                   /number= "7"
FT   intron          62439..73173
FT                   /*tag=  o
FT                   /number= "7"
FT   exon            73174..73248
FT                   /*tag=  p
FT                   /number= "8"
FT   intron          73249..74633
FT                   /*tag=  q
FT                   /number= "8"
FT   exon            74634..74705
FT                   /*tag=  r
FT                   /number= "9"
FT   intron          74706..82788
FT                   /*tag=  s
FT                   /number= "9"
FT   exon            82789..82864
FT                   /*tag=  t
FT                   /number= "10"
FT   intron          82865..83102
FT                   /*tag=  u
FT                   /number= "10"
FT   exon            83103..83194
FT                   /*tag=  v
FT                   /number= "11"
FT   intron          83195..83334
FT                   /*tag=  w
FT                   /number= "11"
FT   exon            83335..83414
FT                   /*tag=  x
FT                   /number= "12"
FT   intron          83415..88129
FT                   /*tag=  y
FT                   /number= "12"
FT   exon            88130..88172
FT                   /*tag=  z
FT                   /number= "13"
FT   intron          88173..89561
FT                   /*tag=  aa
FT                   /number= "13"
FT   exon            89562..89641
FT                   /*tag=  ab
FT                   /number= "14"
FT   intron          89642..91162
FT                   /*tag=  ac
FT                   /number= "14"
FT   exon            91163..91233
FT                   /*tag=  ad
FT                   /number= "15"
FT   intron          91234..93443
FT                   /*tag=  ae
FT                   /number= "15"
FT   exon            93444..93484
FT                   /*tag=  af
FT                   /number= "16"
FT   intron          93485..100599
FT                   /*tag=  ag
FT                   /number= "16"
FT   exon            100600..102009
FT                   /*tag=  ah
FT                   /number= "17"
XX
PN   FR2798136 A1.
XX
PD   09-MAR-2001.
XX
PF   03-SEP-1999;   99FR-0011097.
XX
PR   03-SEP-1999;   99FR-0011097.
XX
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Weissenbach J,  Hazan J;
XX
DR   WPI; 2001-283966/30.
XX
PT   New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT   autosomal dominant familial spastic paraplegia and in drug screening
XX
PS   Claim 2; Page 45-106; 145pp; French.
XX
CC   The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC   a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC   autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC   and their fragments, are used to screen DNA banks for sequences that
CC   encode spastin (particularly sequences in other mammals, specifically
CC   mice); to identify SPG4 mutations, or other genetic anomalies,
CC   particularly for diagnosis of autosomal dominant familial spastic
CC   paraplegia (PSF-AD); to identify promoters and other regulatory elements
CC   of the SPG4 gene; for detection and amplification; for recombinant
CC   production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ   Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 0 other;

  Query Match             100.0%;  Score 501;  DB 22;  Length 110000;
  Best Local Similarity   100.0%;  Pred. No. 2.8e-125;
  Matches  501;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50000 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 50059

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50060 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 50119

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50120 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 50179

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50180 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 50239

Qy        241 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50240 GGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTA 50299

Qy        301 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      50300 TTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTG 50359

Qy        361 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 420

Db      50360 CTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCT 50419

Qy        421 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 480

Db      50420 TTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTT 50479

Qy        481 TATCAGAGAGTTGGATTGCAA 501

Db      50480 TATCAGAGAGTTGGATTGCAA 50500


RESULT 2
AAC69111
ID   AAC69111 standard; DNA; 5065 BP.
XX
AC   AAC69111;
XX
DT   31-JAN-2001  (first entry)
XX
DE   Human secreted protein gene 3 clone HPCCX13.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055371-A1.
XX
PD   21-SEP-2000.
XX
PF   16-MAR-2000; 2000WO-US06783.
XX
PR   18-MAR-1999;  99US-0125059.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI   Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX
DR   WPI; 2000-594448/56.
DR   P-PSDB; AAB38011.
XX
PT   New nucleic acid molecules encoding 27 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions and
PT   used as food additives or preservatives -
XX
PS   Claim 1; Page 383-384; 453pp; English.
XX
CC   The invention relate to the isolation of genes AAC69084-C69119 encoding
CC   27 human secreted proteins AAB37984-B38019. The genes can be used to
CC   generate fusion proteins by linking to the gene for the human
CC   immunoglobulin G Fc portion (AAC69075) for increasing the stability of
CC   the fusion protein as compared to the human protein only. The genes and
CC   proteins are useful for preventing, ameliorating or treating medical
CC   conditions, e.g. by protein or gene therapy. The genes are isolated
CC   from a range of human tissues disclosed in the specification. The
CC   nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC   wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
XX
SQ   Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other;

  Query Match             87.0%;  Score 436;  DB 21;  Length 5065;
  Best Local Similarity   91.7%;  Pred. No. 4.4e-108;
  Matches  473;  Conservative    4;  Mismatches   24;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db       3355 GGTTTTGAGTGAGWTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAY 3414

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db       3415 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAATATGT 3474

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db       3475 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 3534

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db       3535 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTGCT 3594

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db       3595 GGGTATCCTTGTTRACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 3654

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db       3655 AGTCTCCCATTATTAWTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 3714

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db       3715 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 3774

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db       3775 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 3834

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db       3835 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 3870


RESULT 3
AAF33219
ID   AAF33219 standard; cDNA; 5065 BP.
XX
AC   AAF33219;
XX
DT   23-MAR-2001  (first entry)
XX
DE   Human secreted protein gene 7 SEQ ID NO:17.
XX
KW   Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW   dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW   immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW   ophthalmological; neuroprotectant; nootropic; anticonvulsant; vulnerary;
KW   antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
KW   multiple sclerosis; systemic lupus erythematosus; HIV; infection;
KW   hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
KW   cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
KW   coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
KW   corneal graft neovascularisation; neurological disorder; regeneration;
KW   Huntington's chorea; Alzheimer's disease; Parkinson's disease;
KW   infectious disease; chemotaxis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200076530-A1.
XX
PD   21-DEC-2000.
```

```
XX
PF   01-JUN-2000; 2000WO-US14933.
XX
PR   11-JUN-1999;   99US-0138572.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Rosen CA,  Ruben SM,  Komatsoulis GA;
XX
DR   WPI; 2001-071147/08.
DR   P-PSDB; AAB64888.
XX
PT   Nucleic acids encoding 49 human secreted polypeptides, useful for
PT   preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT   disease and diabetic retinopathy -
XX
PS   Claim 1; Page 453-455; 554pp; English.
XX
CC   The polynucleotide sequences given in AAF33213 to AAF33261 encode the
CC   human secreted proteins given in AAB64882 to AAB64930. AAB64931 to
CC   AAB64991 represent human secreted polypeptide sequences and proteins
CC   homologous to them, which are given in the exemplification of the present
CC   invention. Human secreted proteins have activities based on the tissues
CC   and cells the genes are expressed in. Examples of activities include:
CC   immunomodulatory; antisclerotic; dermatological; immunosuppressive;
CC   antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;
CC   vascular; antimicrobial; anti-angiogenic; ophthalmological;
CC   neuroprotectant; anticonvulsant; nootropic; antialzheimers;
CC   antiparkinsonian; and vulnerary. The polynucleotides and polypeptides can
CC   be used in the prevention, diagnosis and treatment of diseases associated
CC   with inappropriate polypeptide expression. Disorders that may be
CC   prevented, diagnosed and/or created by the above methods include immune
CC   disorders (e.g. multiple sclerosis, systemic lupus erythematosus and
CC   human immuno-deficiency virus (HIV) infections), hyperproliferative
CC   disorders (e.g. cancers and Gaucher's disease), cardiovascular diseases
CC   (e.g. Scimitar syndrome, Chaga's cardiomyopathy and coronary
CC   arteriosclerosis), angiogenic disorders (e.g. corneal graft
CC   neovascularisation and diabetic retinopathy), neurological disorders
CC   (e.g. Huntington's chorea, Alzheimer's disease and Parkinson's disease),
CC   infectious diseases and/or for promoting wound healing, regeneration and
CC   /or chemotaxis. AAF33204 to AAF33212 and AAB64881 represent sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other;

  Query Match             87.0%;  Score 436;  DB 22;  Length 5065;
  Best Local Similarity   91.7%;  Pred. No. 4.4e-108;
  Matches  473;  Conservative    4;  Mismatches   24;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db       3355 GGTTTTGAGTGAGWTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAY 3414

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db       3415 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAASTATGT 3474

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db       3475 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 3534

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db       3535 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 3594

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db       3595 GGGTATCCTTGTTRACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 3654

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db       3655 AGTCTCCCATTATTAWTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 3714

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db       3715 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 3774

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db       3775 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 3834

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db       3835 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 3870


RESULT 4
ABX08336
ID   ABX08336 standard; DNA; 1691140 BP.
XX
AC   ABX08336;
XX
DT   20-JAN-2003  (first entry)
XX
DE   Human phosphodiesterase 4D (PDE4D) gene.
XX
KW   Human; Phosphodiesterase 4D; PDE4D; stroke; cerebroprotective;
KW   gene; ds; single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       732790
FT                   /*tag=  a
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       735966
FT                   /*tag=  b
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       736226
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       736516
FT                   /*tag=  d
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       850001
FT                   /*tag=  e
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       852776
FT                   /*tag=  f
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       853079
FT                   /*tag=  g
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       853575
FT                   /*tag=  h
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       856468
FT                   /*tag=  i
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       860845
FT                   /*tag=  j
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       870924
FT                   /*tag=  k
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027267
FT                   /*tag=  l
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027643
FT                   /*tag=  m
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1027757
FT                   /*tag=  n
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1028146
```

```
FT                   /*tag= o
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1037657
FT                   /*tag= p
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1044016
FT                   /*tag= q
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1044045
FT                   /*tag= r
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1254737
FT                   /*tag= s
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1254849
FT                   /*tag= t
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1255763
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1257206
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1258161
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1268207
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1267187
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1268553
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1272669
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1272910
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273023
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273220
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273240
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1273543
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1288439
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1289730
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1290176
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1293745
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1344605
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1344864
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1345135
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1345286
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1346112
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1352976
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354291
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354377
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354554
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1354675
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1355114
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1355693
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1357081
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1362985
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363021
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363827
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1363911
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1364061
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1364066
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1367904
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1368193
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1368217
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1373349
FT                   /*tag= bg
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1373384
FT                   /*tag= bh
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1373415
FT                   /*tag= bi
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       1373979
FT                   /*tag= bj
FT                   /standard_name= "Single nucleotide polymorphism"
FT

  Query Match             97.0%;  Score 436;  DB 24;  Length 1691380;
  Best Local Similarity   92.2%;  Pred. No. 2.5e-107;
```

```
  Matches  476;  Conservative    0;  Mismatches   25;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
              ||||||  ||||  ||||  |||  .||  ||  ||| ||  ||    |  | |
Db      12249 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTAAGAGAC 12308

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              |||||  |||| | |||  |||| |||  |||| ||  |||  ||  ||| ||
Db      12309 AGTTTGTCATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 12368

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              | || |||  |||  | |  ||||  |||  |||  |||  |||  ||  |   |||
Db      12369 GGTCAGTTTTGGAATAGGAGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGCTTTGG 12428

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              ||   |||||  ||||  |||||  ||||  |||  |||  |||  ||  |||   |  ||
Db      12429 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 12488

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
              |||                ||||  |||  |||  |||  |  |  |||  ||  ||  ||
Db      12489 GGATATCCTTGTTAACTTTCTGTCATGTGGATCTGTCTAATGTTGACAGTGGGGTGTTGA 12548

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
              |||  ||||  ||  |||||  |||  |||  |  |||  |||  |||  |  |||  |
Db      12549 AGTCTCCCATTATTATTGTGTGGGAGTCTACGTCTCTTAGTAGGTCTCTAAGGACTTGCT 12608

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
              |||||||  |||  |||  |||||  ||||  |||  |||  |||  |||  |||  |
Db      12609 TTATGAATCTGGCTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 12668

Qy        406 GTTGAATTGATCCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
              ||||||  |||||  |||  ||  ||  |  |||||  |||  |||  |  |  ||  |
Db      12669 GTTGAATTGATCCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGACCTTTGTTG 12728

Qy        466 GTTTAAAGTCTGTTTTTATCAGAGAGTTGGATTGCAA 501
              |||||||||  ||||  |  ||||  |  ||  |||  |
Db      12729 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 12764


RESULT 5
AAL35853
ID   AAL35853 standard; DNA; 9192 BP.
XX
AC   AAL35853;
XX
DT   08-JAN-2002  (first entry)
XX
DE   Human musculoskeletal system related polynucleotide SEQ ID NO 2216.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein;
KW   musculoskeletal system; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155367-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01338.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190876.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   26-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225212.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227809.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
```

```
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-451937/48.
XX
PT   Isolated polypeptide for treating, preventing and/ or prognosing
PT   disorders related to the musculoskeletal system including
PT   musculoskeletal cancers and also for testing and detection e.g.
PT   diagnosis -
XX
PS   Example 2; SEQ ID NO 2218; 781pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (AAL34669-AAL37666) and proteins
CC   (ABB03087-ABB04109) associated with the musculoskeletal system useful
CC   for preventing, treating or ameliorating medical conditions e.g. by
CC   protein or gene therapy. The genes are isolated from a range of human
CC   tissues disclosed in the specification. The nucleic acids, proteins,
CC   antibodies and (ant)agonists are useful in the diagnosis, treatment
CC   and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC   other cancers of the adrenal gland, bone, bone marrow, breast,
CC   gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC   disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC   anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC   multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC   (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC   healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC   and (f) infectious diseases such as viral, bacterial, fungal and
CC   parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9192 BP; 2180 A; 1723 C; 1845 G; 3444 T; 0 other;

  Query Match             86.7%;  Score 434.4;  DB 22;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 1.4e-107;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377


RESULT 6
ABK42359
ID   ABK42359 standard; DNA; 9192 BP.
XX
AC   ABK42359;
XX
DT   21-MAY-2002  (first entry)
XX
DE   Genomic sequence #258 encoding novel human connective tissue polypeptide.
XX
KW   Human; connective tissue related disorder; cancer; gene therapy;
KW   cytostatic; gene; ds.
XX
OS   Homo sapiens.
XX
```

```
PN   WO200155343-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01322.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.
PR   22-AUG-2000; 2000US-0227182.
PR   23-AUG-2000; 2000US-0227009.
PR   30-AUG-2000; 2000US-0228924.
PR   01-SEP-2000; 2000US-0229287.
PR   01-SEP-2000; 2000US-0229343.
PR   01-SEP-2000; 2000US-0229344.
PR   01-SEP-2000; 2000US-0229345.
PR   05-SEP-2000; 2000US-0229509.
PR   05-SEP-2000; 2000US-0229513.
PR   06-SEP-2000; 2000US-0230437.
PR   06-SEP-2000; 2000US-0230438.
PR   08-SEP-2000; 2000US-0231242.
PR   08-SEP-2000; 2000US-0231243.
PR   08-SEP-2000; 2000US-0231244.
PR   08-SEP-2000; 2000US-0231413.
PR   08-SEP-2000; 2000US-0231414.
PR   08-SEP-2000; 2000US-0232080.
PR   08-SEP-2000; 2000US-0232081.
PR   12-SEP-2000; 2000US-0231968.
PR   14-SEP-2000; 2000US-0232397.
PR   14-SEP-2000; 2000US-0232398.
PR   14-SEP-2000; 2000US-0232399.
PR   14-SEP-2000; 2000US-0232400.
PR   14-SEP-2000; 2000US-0232401.
PR   14-SEP-2000; 2000US-0233063.
PR   14-SEP-2000; 2000US-0233064.
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   21-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   08-NOV-2000; 2000US-0246474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249218.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259676.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
```

```
XX
DR   WPI; 2001-565190/63.
XX
PT   Nucleic acid encoding novel connective tissue associated polypeptides,
PT   used in diagnosing, preventing, treating or ameliorating a disorder
PT   such as cancer or rheumatoid arthritis -
XX
PS   Disclosure; SEQ ID No 1246; 673pp; English.
XX
CC   The present invention relates to the isolation of novel human connective
CC   tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC   (cDNA and genomic) sequences encoding them. The sequences of the
CC   invention are useful in the diagnosis, treatment, prevention and/or
CC   prognosis of diseases associated with connective tissue(s), including
CC   cancer. The polynucleotide sequences of the invention are also useful
CC   in gene therapy. ABK42102-ABK43116 represent genomic sequences encoding
CC   the novel human connective tissue related polypeptides.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 9192 BP; 2180 A; 1723 C; 1845 G; 3444 T; 0 other;

  Query Match             86.7%;  Score 434.4;  DB 23;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 1.4e-107;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   16;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
                ||||| |||| |||||||||||||||||||||||||||||||||||||| |||||||||
Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
              ||||||| ||||||||||||||||||||||||||||||| ||||| ||||||| ||||||
Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
              ||||||||||||||||||||||||||||||||||| || |||||||||||||||||||||
Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
              |||  ||||||||||||||||||||||||| | |||||||||||||||||||| | ||||
Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
              |||               ||||| ||||||||||||||| |||||||||||||||| 
Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
               | ||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||
Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
              ||||||||||||||||| ||||||||| ||||||||||||||||||||||||||||||||
Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
              |||||||||||||||||||||||| | |||||||||
Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377


ESULT 7
BX58841
D   ABX58841 standard; cDNA; 9192 BP.
X
C   ABX58841;
X
T   26-FEB-2003  (first entry)
X
E   cDNA encoding novel human musculoskeletal system antigen #1185.
XX
KW   Gene; ss; musculoskeletal system antigen; cancer; metastasis;
KW   re-vascularisation; thrombosis; arteriosclerosis; mineral content;
KW   cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;
KW   post-operative tissue repair; limb regeneration; neuronal growth;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   AIDS-related complex; chondrocyte growth; bone regeneration;
KW   periodontal regeneration; tissue transport; bone graft; skin aging;
KW   keratinocyte growth; hair loss; melanocyte growth; cell proliferation;
KW   cell growth; organ transplant; cell differentiation; body height;
KW   weight; hair colour; eye colour; skin; percentage of adipose tissue;
KW   pigmentation; cosmetic surgery; metabolism; biorhythm; caricadic rhythm;
KW   depression; tendency for violence; pain; reproductive capability;
KW   hormone level; endocrine level; appetite; libido; memory; stress;
KW   storage capability; fat content; lipid content; protein content;
KW   carbohydrate content; vitamin content; cofactor content;
KW   nutritional component.
XX
OS   Homo sapiens.
XX
PN   US2002147140-A1.
XX
PD   10-OCT-2002.
XX
PF   17-JAN-2001; 2001US-0764877.
XX
PR   31-JAN-2000; 2000US-179065P.
PR   04-FEB-2000; 2000US-180628P.
PR   28-JUN-2000; 2000US-214886P.
PR   07-JUL-2000; 2000US-216647P.
PR   07-JUL-2000; 2000US-216880P.
PR   11-JUL-2000; 2000US-217487P.
PR   11-JUL-2000; 2000US-217496P.
PR   14-JUL-2000; 2000US-218290P.
PR   26-JUL-2000; 2000US-220963P.
PR   26-JUL-2000; 2000US-220964P.
PR   14-AUG-2000; 2000US-224518P.
PR   14-AUG-2000; 2000US-224519P.
PR   14-AUG-2000; 2000US-225267P.
PR   14-AUG-2000; 2000US-225268P.
PR   14-AUG-2000; 2000US-225270P.
PR   14-AUG-2000; 2000US-225447P.
PR   14-AUG-2000; 2000US-225757P.
PR   14-AUG-2000; 2000US-225758P.
PR   22-AUG-2000; 2000US-226868P.
PR   30-AUG-2000; 2000US-228924P.
PR   01-SEP-2000; 2000US-229287P.
PR   01-SEP-2000; 2000US-229343P.
PR   01-SEP-2000; 2000US-229344P.
PR   01-SEP-2000; 2000US-229345P.
PR   05-SEP-2000; 2000US-229509P.
PR   05-SEP-2000; 2000US-229513P.
PR   08-SEP-2000; 2000US-231413P.
PR   21-SEP-2000; 2000US-234223P.
PR   21-SEP-2000; 2000US-234274P.
PR   25-SEP-2000; 2000US-234997P.
PR   27-SEP-2000; 2000US-235834P.
PR   29-SEP-2000; 2000US-236327P.
PR   29-SEP-2000; 2000US-236367P.
PR   29-SEP-2000; 2000US-236368P.
PR   29-SEP-2000; 2000US-236369P.
PR   29-SEP-2000; 2000US-236370P.
PR   02-OCT-2000; 2000US-236802P.
PR   02-OCT-2000; 2000US-237037P.
PR   02-OCT-2000; 2000US-237038P.
PR   02-OCT-2000; 2000US-237039P.
PR   02-OCT-2000; 2000US-237040P.
PR   13-OCT-2000; 2000US-239935P.
PR   20-OCT-2000; 2000US-240960P.
PR   20-OCT-2000; 2000US-241785P.
PR   20-OCT-2000; 2000US-241809P.
PR   01-NOV-2000; 2000US-244617P.
PR   17-NOV-2000; 2000US-249299P.
```

```
PR   08-DEC-2000; 2000US-251856P.
PR   08-DEC-2000; 2000US-251868P.
PR   08-DEC-2000; 2000US-251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA,  Ruben SM,  Barash SC;
XX
DR   WPI; 2003-126199/12.
XX
PT   Isolated nucleic acid molecules encoding musculoskeletal system
PT   associated polypeptides, useful for detecting disorders, e.g. cancer -
XX
PS   Disclosure; SEQ ID NO 2218; 321pp; English.
XX
CC   The invention describes an isolated nucleic acid molecule comprising a
CC   sequence encoding musculoskeletal system associated polypeptides useful
CC   for detecting disorders, e.g., cancer or cancer metastases, in animals
CC   or humans. The nucleic acid: stimulates re-vascularisation of ischaemic
CC   tissues associated with conditions such as thrombosis, arteriosclerosis,
CC   and other cardiovascular conditions; treats wounds due to injuries,
CC   burns, post-operative tissue repair, and ulcers; stimulates angiogenesis
CC   and limb regeneration; stimulates neuronal growth; can treat and prevent
CC   neuronal damage occurring in certain disorders or neurodegenerative
CC   conditions, such as, Alzheimer's disease, Parkinson's disease, and
CC   AIDS-related complex; stimulates chondrocyte growth, thus they can be
CC   used to enhance bone and periodontal regeneration and aid in tissue
CC   transports or bone grafts; prevents skin aging due to sunburn by
CC   stimulating keratinocyte growth; prevents hair loss, since FGF family
CC   members activate hair-forming cells and promotes melanocyte growth;
CC   stimulates growth and differentiation of hematopoietic cells and bone
CC   marrow cells when used in combination with other cytokines; maintains
CC   organs before transplantation or for supporting cell culture of primary
CC   tissues; induces tissue of mesodermal origin to differentiate in early
CC   embryos; increases or decreases the differentiation or proliferation of
CC   embryonic stem cells, besides, haematopoietic lineage; modulates
CC   mammalian characteristics, such as, body height, weight, hair colour, eye
CC   colour, skin, percentage of adipose tissue, pigmentation, size, and shape
CC   (e.g., cosmetic surgery); modulates mammalian metabolism; changes
CC   mammal's metal state or physical state by influencing biorhythms,
CC   caricadic rhythms, depression, tendency for violence, tolerance for pain,
CC   reproductive capabilities, hormonal or endocrine levels, appetite,
CC   libido, memory, or stress; increases or decreases storage capabilities,
CC   fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors
CC   or other nutritional components. This sequence encodes a novel human
CC   musculoskeletal system antigen.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from the US patent office at
CC   ftp.seqdata.uspto.gov/sequence.html?DocID=20020147140.
XX
SQ   Sequence 9192 BP; 2180 A; 1723 C; 1845 G; 3444 T; 0 other;

  Query Match             86.7%;  Score 434.4;  DB 25;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 1.4e-107;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377


RESULT 8
ABK92491/c
ID   ABK92491 standard; DNA; 1865 BP.
XX
AC   ABK92491;
XX
DT   19-AUG-2002  (first entry)
XX
DE   Human prostate specific nucleic acid #110.
XX
KW   Prostate specific polypeptide; metastasis; prostate cancer; cancer;
KW   non-cancerous prostate disease; gene therapy; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200238810-A2.
XX
PD   16-MAY-2002.
XX
PF   06-NOV-2001; 2001WO-US47001.
XX
PR   06-NOV-2000; 2000US-246109P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Sun Y,  Recipon H,  Chen S,  Liu C;
XX
DR   WPI; 2002-471559/50.
XX
PT   New prostate specific polypeptides and polynucleotides useful for
PT   detecting, diagnosing, monitoring, treating, staging and predicting
PT   cancers in humans having cancer and non-cancerous prostate disease  -
XX
PS   Claim 1; Page 206-207; 267pp; English.
XX
CC   The invention describes an isolated prostate specific polypeptide (I) and
CC   nucleic acid (II) encoding it and are useful for diagnosing and
CC   monitoring the presence and metastases of prostate cancer in a patient.
CC   (I), (II) and an antibody to (II) are useful in quantitative and
CC   qualitative diagnostic assays and methods for detecting, diagnosing,
CC   monitoring, treating, staging and predicting cancers in humans having
CC   cancer or may have the risk of developing cancer. (I) and (II) are also
CC   useful for determining non-cancerous prostate disease, by measuring
CC   their expression levels and/or structural alterations; for determining
CC   the sample that has prostate tissue-like characteristics or is a prostate
CC   tissue; as an element in an array or a multi-analyte test to recognise
CC   expression patterns associated with prostate cancer and other prostate
CC   related disorders; and as elements in a computer program for pattern
CC   recognition of prostate disorders. (I) and (II) are useful for producing
CC   engineered prostate tissue for treatment and research. (II) is useful for
```

```
CC   producing transgenic animals and cells and also in gene therapy. This
CC   sequence represents a prostate specific nucleic acid described in
CC   the invention.
XX
SQ   Sequence 1885 BP; 662 A; 421 C; 398 G; 404 T; 0 other;

  Query Match             86.4%;  Score 432.8;  DB 24;  Length 1885;
  Best Local Similarity   91.9%;  Pred. No. 2.4e-107;
  Matches  474;  Conservative    0;  Mismatches   27;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        733 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGCGGTCTGAGAGAC 674

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        673 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 614

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        613 GGTCAATTTTGGAATAAGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 554

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        553 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 494

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        493 GGATATCCTTGTTAACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTAGGGTGTTAA 434

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        433 AGTCTCCCACTATTATTGTGTGGGAGTTTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 374

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        373 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTCAGGATAGTTAGCTCTTCTT 314

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        313 GTTGAATTGATCCCTTTACCACTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTT 254

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        253 GTTTAACGTCTGTTTTATCAGAGACTAGGATTGCAA 218


RESULT 9
AAZ35351/c
ID   AAZ35351 standard; DNA; 41599 BP.
XX
AC   AAZ35351;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Cosmid including sequence spanning human chromosome 9p21.
XX
KW   Chromosome 9p21; human; cancer; tumour; diagnosis; prognosis;
KW   methylthioadenosine phosphorylase; glioma; melanoma;
KW   primary lymphoid malignancy; non-small cell lung cancer;
KW   head and neck cancer; ovarian cancer; bladder cancer;
KW   chondrosarcoma; ds.
XX
OS   Homo sapiens.
XX
PN   WO9967634-A1.
XX
PD   29-DEC-1999.
XX
PF   21-JUN-1999;   99WO-US13981.
XX
PR   23-JUN-1998;   98US-0090411.
PR   17-JUN-1999;   99US-0335231.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Carson DA,  Schmid M,  Carrera CJ;
XX
DR   WPI; 2000-126650/11.
XX
PT   Diagnosing and determining prognosis for cancer causitively associated
PT   with derangements of chromosome 9p21  -
XX
PS   Disclosure; Fig 7; 55pp; English.
XX
CC   This is the nucleotide sequence of a cosmid that includes sequences
CC   spanning human chromosome 9p21 (ATCC AC00047).  This region
CC   harbours 5 genes within about 120 kb: the tumour suppressor genes
CC   p15INK4B (p15) with its alternatively spliced form p10, p16INK4A
CC   (p16) and p19ARF, and the gene for methylthioadenosine phosphorylase
CC   (MTAP).  The invention provides a method for diagnosis of, and
CC   determining a prognosis for, cancer causatively associated with
CC   derangements of chromosome 9p21.  Underlying the invention is the
CC   discovery that such derangements have their genesis in deletions
CC   occurring centromeric to STS 3.21, most often including breakpoints
CC   in exon 6 and/or between exons 4 and 5 of the gene which encodes
CC   MTAP.  As the cancer and tumour development advance, deletions in
CC   9p21 progress centromerically from the genesis point toward the
CC   gene encoding p16.  Thus, the method of the invention is performed
CC   by determining whether (a) portions of the 9p21 region including
CC   and telomeric to STS 3.21 are deleted, and (b) portions of the
CC   9p21 region centromeric to SRS 3.21 are deleted, where a positive
CC   finding in step (a) and a negative finding in step (b) are
CC   indicative of a cancer in an early stage of tumour development
CC   and a positive finding in step (a) is indicative of a cancer in an
CC   advanced stage of tumour development.  Primer pairs (see AAZ35354-75)
CC   are provided for use in claimed methods for diagnosing and
CC   determining a prognosis for cancer associated with derangements of
CC   9p21, especially a glioma, primary lymphoid malignancy, non-small
CC   cell lung cancer, melanoma, head and neck cancer, ovarian cancer,
CC   bladder cancer or a chondrosarcoma (claimed).
XX
SQ   Sequence 41599 BP; 13477 A; 8216 C; 8242 G; 11664 T; 0 other;

  Query Match             86.1%;  Score 431.6;  DB 21;  Length 41599;
  Best Local Similarity   91.8%;  Pred. No. 1.3e-106;
  Matches  472;  Conservative    0;  Mismatches   29;  Indels   13;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      11896 GGTTTTGTGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 11837

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      11836 AGTTTATTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 11777

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      11776 GGTCAATTTTGGAATAGGTGCAGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 11717

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      11716 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAATTCAATTCCT 11657

Qy        241 GGA-------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAG 287
Db      11656 GGATATCCTTGTTAACTTTGTCTCATTGATCTGTCTAATGTTGATAGTGGGGTGTTAAAG 11597

Qy        288 TCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTT 347
Db      11596 TCTCCCATTATTACTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCTTT 11537

Qy        348 ATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGT 407
Db      11536 ATGAATCTGAGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGT 11477
```

```
Qy        408 TGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGT 467
Db      11476 TGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGT 11417

Qy        468 TTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      11416 TTAAAGTCTGTTTTATGAGAGACTAGGATTGCAA 11383
```

RESULT 10
AAS84982/c
ID   AAS84982 standard; cDNA; 2277 BP.
XX
AC   AAS84982;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #20786.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG20795.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 20786; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2277 BP; 927 A; 501 C; 411 G; 438 T; 0 other;

```
  Query Match             86.1%;  Score 431.2;  DB 23;  Length 2277;
  Best Local Similarity   91.7%;  Pred. No. 6.9e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       1123 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 1064

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       1063 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 1004

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       1003 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 944

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        943 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 884

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        883 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 824

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        823 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 764

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        763 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 704

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        703 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 644

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        643 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 608
```

RESULT 11
AAS75832/c
ID   AAS75832 standard; cDNA; 2606 BP.
XX
AC   AAS75832;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #11636.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG11645.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess

```
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 11636; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2606 BP; 982 A; 600 C; 478 G; 546 T; 0 other;

  Query Match             86.1%;  Score 431.2;  DB 23;  Length 2606;
  Best Local Similarity   91.7%;  Pred. No. 7.2e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       1443 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 1384

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       1383 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 1324

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       1323 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 1264

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       1263 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 1204

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       1203 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 1144

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       1143 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 1084

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       1083 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 1024

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       1023 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 964

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        963 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 928

ESULT 12
AS70259/c
D   AAS70259 standard; cDNA; 3316 BP.
X
AC   AAS70259;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #6063.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG06072.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 6063; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3316 BP; 1389 A; 713 C; 561 G; 653 T; 0 other;

  Query Match             86.1%;  Score 431.2;  DB 23;  Length 3316;
  Best Local Similarity   91.7%;  Pred. No. 7.8e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        729 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 670

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        669 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 610

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        609 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 550
```

```
Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        549 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 490

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        489 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 430

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        429 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 370

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        369 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 310

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        309 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 250

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        249 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 214
```

RESULT 13
AAS70804/c
ID AAS70804 standard; cDNA; 3316 BP.
XX
AC AAS70804;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #6608.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG06617.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 6608; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 3316 BP; 1383 A; 713 C; 561 G; 653 T; 0 other;

```
  Query Match             86.1%;  Score 431.2;  DB 23;  Length 3316;
  Best Local Similarity   91.7%;  Pred. No. 7.8e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        729 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 670

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        669 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 610

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        609 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 550

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        549 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 490

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        489 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 430

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        429 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 370

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        369 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 310

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        309 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 250

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        249 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 214
```

RESULT 14
AAS73840/c
ID AAS73840 standard; cDNA; 3316 BP.
XX
AC AAS73840;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #9644.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX

```
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG09653.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 9644; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 3316 BP; 1389 A; 713 C; 561 G; 653 T; 0 other;

  Query Match             86.1%;  Score 431.2;  DB 23;  Length 3316;
  Best Local Similarity   91.7%;  Pred. No. 7.8e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        729 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 670

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        669 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 610

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        609 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 550

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        549 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 490

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        489 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 430

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        429 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 370

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        369 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 310

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        309 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 250

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        249 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 214

RESULT 15
AAS75000/c
ID   AAS75000 standard; cDNA; 3316 BP.
XX
AC   AAS75000;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #10804.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG10813.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 10804; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SC   Sequence 3316 BP; 1389 A; 713 C; 561 G; 653 T; 0 other;

  Query Match             96.1%;  Score 431.2;  DB 23;  Length 3316;
  Best Local Similarity   91.7%;  Pred. No. 7.8e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db        729 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCGAGTTTGATTGCACTGTGGTCTGAGAGAC 670

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db        669 AGTTTGTTATAATTTCTGTTGTTTTATATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 610

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db        609 GGTCAATTTTGGAATAAATGTGGTGTAGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 550

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db        549 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAAATCCT 490

Qy        241 GGA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db        489 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 430

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db        429 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 370

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db        369 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGGTAGTTAGCTCTTCTT 310

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db        309 GTTGAATTGATCCCTTTACCATCATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 250

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db        249 GTTTAAAGTCCGTTTTATCAGAGACTAGGATTGCAA 214


Search completed: October 24, 2003, 18:47:22
Job time : 131.251 secs
```

```
M nucleic - nucleic search, using sw model

Run on:         October 24, 2003, 13:34:49 ; Search time 31.3542 Seconds
                                             (without alignments)
                                             7052.739 Million cell updates/sec

itle:           US-09-830-902-1_COPY_50000_50500
Perfect score:  501
Sequence:       1 ggttttgcgtgagtttctta..........atcagagagttggattgcaa 501

coring table:   IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

earched:        569978 seqs, 220691566 residues

otal number of hits satisfying chosen parameters:      1139356

inimum DB seq length: 0
aximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
                Maximum Match 100%
                Listing first 45 summaries

atabase :        Issued_Patents_NA:*
                 1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                 2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                 3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                 4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                 5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                 6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	esult No.	Score	% Query Match	Length	DB	ID	Description
	1	428	85.4	116592	4	US-09-818-512-3	Sequence 3, Appli
	2	425.4	84.9	168575	4	US-09-426-290-1	Sequence 1, Appli
	3	424.8	84.8	32042	4	US-09-245-281-44	Sequence 44, Appl
	4	424.8	84.8	32042	4	US-09-340-620A-63	Sequence 63, Appl
	5	412.2	82.3	40328	3	US-08-742-185-102	Sequence 102, App
	6	405.2	80.9	202001	4	US-09-734-674-3	Sequence 3, Appli
	7	361.4	72.1	468	4	US-09-702-705-604	Sequence 604, App
	8	361.4	72.1	468	4	US-09-736-457-604	Sequence 604, App
c	9	358	71.5	168575	4	US-09-426-290-1	Sequence 1, Appli
	10	272.6	54.4	668	3	US-09-385-982-157	Sequence 157, App
	11	230	45.9	18596	3	US-09-318-448-11	Sequence 11, Appl
	12	230	45.9	18597	4	US-09-962-665-8	Sequence 8, Appli
c	13	192.2	38.4	3033	4	US-09-525-160B-9	Sequence 9, Appli
c	14	192.2	38.4	3213	4	US-09-525-160B-4	Sequence 4, Appli
c	15	187.4	37.4	9573	4	US-09-220-132-168	Sequence 168, App
c	16	184.8	36.9	1001	4	US-09-641-638-218	Sequence 218, App
	17	184.8	36.9	29604	3	US-08-781-891-207	Sequence 207, App
	18	184.8	36.9	29604	4	US-09-618-166-207	Sequence 207, App
c	19	184.8	36.9	148567	4	US-09-801-876B-3	Sequence 3, Appli
	20	172.8	34.5	38564	4	US-09-734-673-3	Sequence 3, Appli
c	21	156	31.1	116592	4	US-09-818-512-3	Sequence 3, Appli
	22	154.8	30.9	70000	4	US-09-851-896-3	Sequence 3, Appli
	23	145.4	29.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
	24	145.4	29.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
	25	145.4	29.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	26	140.6	28.1	51259	3	US-08-781-891-209	Sequence 209, App
c	27	140.6	28.1	51259	4	US-09-618-166-209	Sequence 209, App
	28	139.8	27.9	51259	3	US-08-781-891-209	Sequence 209, App
	29	139.8	27.9	51259	4	US-09-618-166-209	Sequence 209, App
c	30	134.4	26.8	64467	4	US-09-803-671B-3	Sequence 3, Appli
	31	133.2	26.6	50000	4	US-09-146-053-4	Sequence 4, Appli
	32	122.4	24.4	1540	4	US-09-393-634-8	Sequence 8, Appli
	33	121.2	24.2	112132	4	US-09-741-150-3	Sequence 3, Appli
c	34	115.4	23.0	648	4	US-09-702-705-609	Sequence 609, App
c	35	115.4	23.0	648	4	US-09-736-457-609	Sequence 609, App
	36	110	22.0	43795	3	US-08-742-185-101	Sequence 101, App
c	37	106.2	21.2	202001	4	US-09-734-674-3	Sequence 3, Appli
c	38	99.8	19.9	623	3	US-09-385-982-167	Sequence 167, App
c	39	96.6	19.3	15418	4	US-09-783-203-1	Sequence 1, Appli
	40	95.4	19.0	319608	4	US-09-539-333D-1	Sequence 1, Appli
	41	95.4	19.0	319608	4	US-09-679-409-1	Sequence 1, Appli
	42	94.6	18.9	49136	3	US-09-422-869-1	Sequence 1, Appli
c	43	90.2	18.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	44	90.2	18.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	45	90.2	18.0	246240	2	US-08-724-394A-22	Sequence 22, Appl

ALIGNMENTS

```
RESULT 1
US-09-818-512-3
; Sequence 3, Application US/09818512
; Patent No. 6537780
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001192
;  CURRENT APPLICATION NUMBER: US/09/818,512
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 116592
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(116592)
;   OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

  Query Match             85.4%;  Score 428;  DB 4;  Length 116592;
  Best Local Similarity   91.3%;  Pred. No. 7.4e-114;
  Matches  471;  Conservative    0;  Mismatches   30;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db      51828 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 51887

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db      51888 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 51947

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db      51948 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 52007

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db      52008 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 52067

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db      52068 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTTAA 52127

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
```

```
Db     52128 AGTCTTCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 52187

Qy       346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     52188 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 52247

Qy       406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db     52248 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 52307

Qy       466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db     52308 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 52343
```

RESULT 2
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-09-426-290-1

```
  Query Match             84.9%;  Score 425.4;  DB 4;  Length 168575;
  Best Local Similarity   91.1%;  Pred. No. 4.7e-113;
  Matches  469;  Conservative    0;  Mismatches   31;  Indels   15;  Gaps    1;

Qy         2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db    156093 GTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACCGTGGTCTGAGAGACA 156152

Qy        62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db    156153 GTTTGTTATAATATCTGATCTTATACATTTGCTGAGGAGAGCTTTACTTCCAACTATGTG 156212

Qy       122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db    156213 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTCGG 156272

Qy       182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db    156273 GTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCTG 156332

Qy       242 GA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db    156333 GATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTTATGTTGACAGTGGGGTGTTAAA 156392

Qy       287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db    156393 GTCTCCCATTATTATTGTGTGGTAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCTT 156452

Qy       347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db    156453 TATGAATCTGGGTGCTCCTATATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 156512

Qy       407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGG 466
Db    156513 TTCAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGG 156572

Qy       467 TTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db    156573 TTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 156607
```

RESULT 3
US-09-245-281-44
; Sequence 44, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 32042
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-245-281-44

```
  Query Match             84.8%;  Score 424.8;  DB 4;  Length 32042;
  Best Local Similarity   90.9%;  Pred. No. 3.9e-113;
  Matches  469;  Conservative    0;  Mismatches   32;  Indels   15;  Gaps    1;

Qy         1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db     15442 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 15501

Qy        61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db     15502 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 15561

Qy       121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db     15562 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 15621

Qy       181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db     15622 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 15681

Qy       241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db     15682 GGGTATCCTTGTTGACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 15741

Qy       286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db     15742 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 15801

Qy       346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     15802 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 15861
```

```
Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      15862 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 15921

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      15922 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 15957


RESULT 4
US-09-340-620A-63
  Sequence 63, Application US/09340620A
  Patent No. 6482933
, GENERAL INFORMATION:
   APPLICANT: Bertin, John
,  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
,  FILE REFERENCE: 07334-124001
   CURRENT APPLICATION NUMBER: US/09/340,620A
,  CURRENT FILING DATE:  1999-06-28
   PRIOR APPLICATION NUMBER: US 09/245,281
   PRIOR FILING DATE: 1999-02-05
   PRIOR APPLICATION NUMBER: US 09/207,359
   PRIOR FILING DATE: 1998-12-08
   PRIOR APPLICATION NUMBER: US 09/099,041
   PRIOR FILING DATE: 1998-06-17
   PRIOR APPLICATION NUMBER: US 09/019,942
   PRIOR FILING DATE: 1998-02-06
   NUMBER OF SEQ ID NOS: 71
   SOFTWARE: FastSEQ for Windows Version 4.0
  SEQ ID NO 63
    LENGTH: 32042
    TYPE: DNA
    ORGANISM: Homo sapiens
US-09-340-620A-63

  Query Match             84.8%;  Score 424.6;  DB 4;  Length 32042;
  Best Local Similarity   90.9%;  Pred. No. 3.9e-113;
  Matches  469;  Conservative    0;  Mismatches   38;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      15442 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 15501

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      15502 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 15561

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      15562 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAAATGTATATTCTGTTGATTTGG 15621

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      15622 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 15681

Qy        241 GG--------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      15682 GGGTATCCTTGTTGACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 15741

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      15742 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 15801

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      15802 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 15861

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      15862 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 15921

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      15922 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 15957


RESULT 5
US-08-742-185-102
; Sequence 102, Application US/08742185
; Patent No. 6020476
;  GENERAL INFORMATION:
;    APPLICANT:  Page, David C.
;    APPLICANT:  Reijo, Renee
;    APPLICANT:  Saxena, Richa
;    APPLICANT:  Hawkins,Trevor
;    APPLICANT:  Reeve, Mary Pat
;    TITLE OF INVENTION:  DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
;    NUMBER OF SEQUENCES:  102
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  US
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/742,185
;      FILING DATE:  30-OCT-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/690,734
;      FILING DATE:  31-JUL-1996
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/310,429
;      FILING DATE:  22-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  WHI94-07A2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  102:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  40328 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-742-185-102

  Query Match             82.3%;  Score 412.2;  DB 3;  Length 40328;
  Best Local Similarity   90.5%;  Pred. No. 1.9e-109;
  Matches  468;  Conservative    0;  Mismatches   33;  Indels   16;  Gaps    2;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      19661 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 19720

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      19721 AGTTTGTTATAATTTCTGTTCTTTTACATCCGCTGAGGAGAGCTTTACTTCCCAGTATGT 19780

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTG-AGAAGAATGTATATTCTGTTGATTTG 179
Db      19781 GGTCAGTTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAAATGTATATTCTGTTGATTTG 19840

Qy        180 GGGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCC 239
Db      19841 GGGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTCC 19900
```

```
Qy        240 TGG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTG 284
Db      19901 TGGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTA 19960

Qy        285 AAGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGC 344
Db      19961 AAGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGC 20020

Qy        345 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 404
Db      20021 TTTATGAATCTTGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 20080

Qy        405 TGTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 464
Db      20081 TGTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTCGATCTTTGTT 20140

Qy        465 GGTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      20141 GGTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 20177
```

RESULT 6
US-09-734-674-3
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

```
  Query Match             80.9%;  Score 405.2;  DB 4;  Length 202001;
  Best Local Similarity   89.9%;  Pred. No. 3.5e-107;
  Matches  464;  Conservative    0;  Mismatches   33;  Indels   19;  Gaps    2;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db     145446 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 145505

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db     145506 AGTTTGTTATAATTTCTGATCTTTTACATTTGCTGAGAAGAGCTTTACTTCCAACTATGT 145565

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db     145566 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 145625

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db     145626 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 145685

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db     145686 GGGTATCCTTGTCAACTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGATGTTAA 145745

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db     145746 AGTCTCCCATTATTATTTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 145805

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     145806 TTATGAATCTGGGTGCTCCTGTATTAGATACATATATATTTAGGATAGTTAGCTCTTCTT 145865

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db     145866 GTTGAGTTGATCCCTTTACCATTATGTAATGGCC----TTGTCTCTTTTGATCTTTGTTG 145921

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db     145922 GTTTAAAGTCTGTTTTATCAGAGACTATGATTGCAA 145957
```

RESULT 7
US-09-702-705-604
; Sequence 604, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 604
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(468)
; OTHER INFORMATION: n = A,T,C or G
US-09-702-705-604

```
  Query Match             72.1%;  Score 361.4;  DB 4;  Length 468;
  Best Local Similarity   88.1%;  Pred. No. 1.9e-95;
  Matches  408;  Conservative    0;  Mismatches   40;  Indels   15;  Gaps    1;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db          4 GTTTTGAGTGAGTTTCTTAATCCTGAGTTCTGGNTTGATTGCACTGTGGTCTGAGAGATA 63

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db         64 GTTTGTTATAATTTCTGTTCTTTTACACTTACTGAGGAGAGCTTTACTTCCAAGTATGTG 123

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db        124 GTCGATTTTGGAATAGGTGTGGTGTCGTGCTGAAAAGAATGTATATTCTGTTGATTTGGG 183

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db        184 GTGGAGAGTTCTGTANATGTCTATTAGGTCCGCTTGGTGCAGAGTTGAGTTCAATTCCTG 243

Qy        242 GA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db        244 GATAGCCTTGTTAACTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGGTAAA 303

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db        304 GTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTAAGGACTTGCTT 363
```

```
Qy        347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db        364 TATGAATCTGGGTGCTCCTGCATTGGGTGCACATATATTTAGGACAGCNAGCTCTTCTTG 423

Qy        407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCT 449
Db        424 TTGAATTGATCCCTTTACCATTATGTAATGGCCTTGNCTCTTT 466


RESULT 8
US-09-736-457-604
; Sequence 604, Application US/09736457
  Patent No. 6509448
  GENERAL INFORMATION:
  APPLICANT:  Wang, Tongtong
  APPLICANT:  Bangur, Chaitanya S.
  APPLICANT:  Lodes, Michael A.
  APPLICANT:  Fanger, Gary
  APPLICANT:  Vedvick, Tom
  APPLICANT:  Carter, Darrick
  APPLICANT:  Retter, Marc
  APPLICANT:  Mannion, Jane
  APPLICANT:  Fan, Ligun
  APPLICANT:  Wang, Aijun
  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
  FILE REFERENCE:  210121.478C15
  CURRENT APPLICATION NUMBER: US/09/736,457
  CURRENT FILING DATE:  2000-12-13
  NUMBER OF SEQ ID NOS:  1864
  SOFTWARE:  FastSEQ for Windows Version 3.0
  SEQ ID NO 604
    LENGTH: 468
    TYPE: DNA
    ORGANISM: Homo sapien
    FEATURE:
    NAME/KEY: misc_feature
    LOCATION: (1)...(468)
    OTHER INFORMATION: n = A,T,C or G
S-09-736-457-604

 Query Match           72.1%;  Score 361.4;  DB 4;  Length 468;
 Best Local Similarity  88.1%;  Pred. No. 1.9e-95;
 Matches  408;  Conservative    0;  Mismatches   40;  Indels   15;  Gaps    1;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db          4 GTTTTGAGTGAGTTTCTTAATCCTGAGTTCTGGNTTGATTGCACTGTGGTCTGAGAGATA 63

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db         64 GTTTGTTATAATTTCTGTTCTTTTACACTTACTGAGGAGAGCTTTACTTCCAAGTATGTG 123

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db        124 GTCGATTTTGGAATAGGTGTGGTGTCGTGCTGAAAAGAATGTATATTCTGTTGATTTGGG 183

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db        184 GTGGAGAGTTCTGTANATGTCTATTAGGTCCGCTTGGTGCAGAGTTGAGTTCAATTCCTG 243

Qy        242 GA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db        244 GATAGCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGGTAAA 303

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db        304 GTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTAAGGACTTGCTT 363

Qy        347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db        364 TATGAATCTGGGTGCTCCTGCATTGGGTGCACATATATTTAGGACAGCNAGCTCTTCTTG 423

Qy        407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCT 449
Db        424 TTGAATTGATCCCTTTACCATTATGTAATGGCCTTGNCTCTTT 466


RESULT 9
US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
;  APPLICANT: Berglind Ran Olafsdottir
;  APPLICANT: Jeffrey Gulcher
;  TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
;  FILE REFERENCE: 2345.2001-000
;  CURRENT APPLICATION NUMBER: US/09/426,290
;  CURRENT FILING DATE: 1999-10-25
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 168575
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (21181)...(21403)
;   NAME/KEY: CDS
;   LOCATION: (95252)...(95430)
;   NAME/KEY: CDS
;   LOCATION: (101753)...(101996)
;   NAME/KEY: CDS
;   LOCATION: (110324)...(110439)
;   NAME/KEY: CDS
;   LOCATION: (124056)...(124278)
;   NAME/KEY: CDS
;   LOCATION: (127009)...(127130)
;   NAME/KEY: CDS
;   LOCATION: (128910)...(129139)
US-09-426-290-1

  Query Match           71.5%;  Score 358;  DB 4;  Length 168575;
  Best Local Similarity  84.1%;  Pred. No. 1.5e-93;
  Matches  434;  Conservative    0;  Mismatches   65;  Indels   17;  Gaps    2;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       7761 GGTTTTGAGTGAGTTTCTTATTCCAGAGCTCTTATTTGATTGCACTGTGGTCTGAGAGAC 7722

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       7721 TGTTTGTTATGATTTTAGTTCTTTGGCATTTGCTGAGGAGTCTTTTACTTCCAATGATGT 7662

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       7661 GGTCAATTTTAGAATAAGTGCAATGTGGTGCTGAGAATAATGTATATTCTATTTATTTGG 7602

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7601 GGTGGAGAGTCCTATAGATGTCTATTAGGCCTGCTTGGTCCAGAGCTGAGTTCAAGTCCT 7542

Qy        241 GGA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7541 GAATGTACTTGTTAATTTTCTGTCTCATTGATCTGTTTAATATTGACAGTGGGGTGTTAA 7482

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7481 AGTCTCCCACTATTATTACGTGGGAGTCTAAGTCTC--TATAGGTCTCTAAGAACTTGCT 7424

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       7423 TGATGAATTTGTGTGCTCCTGAATTGGGTGCATATATATTTAGGATAGCTAGCTGTTCTT 7364

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
```

```
Db       7363 GTTGCATTGATCCCTTTATCATTATGTAATGCCCTTTGTTGTCTCTTTTGATCTTTGTTG 7304

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db       7303 GTTTGAAGTCTATTTTATCAGAGACTAGGATTGCAA 7268


RESULT 10
US-09-385-982-157
; Sequence 157, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
;  APPLICANT: ENDEGE, WILSON O., ET AL.
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
;  TITLE OF INVENTION: PRODUCTS: II
;  FILE REFERENCE: CCDNA-260XX
;  CURRENT APPLICATION NUMBER: US/09/385,982
;  CURRENT FILING DATE: 1999-08-30
;  EARLIER APPLICATION NUMBER: 09/328,111
;  EARLIER FILING DATE: 1999-06-08
;  EARLIER APPLICATION NUMBER: 60/117,393
;  EARLIER FILING DATE: 1999-01-27
;  EARLIER APPLICATION NUMBER: 60/098,639
;  EARLIER FILING DATE: 1998-08-31
;  NUMBER OF SEQ ID NOS: 544
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 157
;   LENGTH: 668
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(668)
;   OTHER INFORMATION: n = A,T,C or G
US-09-385-982-157

  Query Match             54.4%;  Score 272.6;  DB 3;  Length 668;
  Best Local Similarity   86.2%;  Pred. No. 1e-69;
  Matches  374;  Conservative    0;  Mismatches   42;  Indels   18;  Gaps    6;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db         53 GGTTTTGAGTGAGTTTCTTAAACCTGAGTTGTCGTTTGATTGCACTGTGGTCTGAGAGAC 112

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db        113 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCACCTATGT 172

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db        173 GGTCAATTTTGGAATAAGTGAGATGTGGTGCTAAAAAGAATATATATTCTGTTGATTTGA 232

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db        233 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCANAGCTGAGTCAATTCCTG 292

Qy        241 G-------------GATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAG 287

Db        293 GATATCCTTGGTAACTTTCTGCTTGTTGNTCTGTCTAATATTGACAGTGGGGCGTTAAAG 352

Qy        288 TCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTT 347

Db        353 TCTCCCA-TATTATTGTGTGGGAGTCTAA-TCTCTTTGTAGGTCTCTAAGGACTTGCTTT 410

Qy        348 ATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGT 407

Db        411 AT-AAACTGGGTGCTCTTG-ATTGGGTGCA-ATATATTTAGGATAGTTAGCTCTTCTTGG 467

Qy        408 TGAATTGATCCCTT 421

Db        468 TGAATGGANCCTTT 481


RESULT 11
US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 18596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-11

  Query Match             45.9%;  Score 230;  DB 3;  Length 18596;
  Best Local Similarity   87.4%;  Pred. No. 6.9e-57;
  Matches  304;  Conservative    0;  Mismatches   25;  Indels   19;  Gaps    4;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61

Db       9059 GTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGATA 9118

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121

Db       9119 GTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGTG 9178

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181

Db       9179 GTCGGTTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGGG 9238

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241

Db       9239 --ATGGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCTG 9296

Qy        242 G--------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286

Db       9297 GGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTGTACTGTTGACAGT-GGGTGTTAAA 9355

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCT 334

Db       9356 GTCTCCCATTATTAATGTGT-GGAGTCTAAGTCTCTTTGTAGGTCACT 9402


RESULT 12
US-09-962-665-8
; Sequence 8, Application US/09962665
; Patent No. 6537759
; GENERAL INFORMATION:
;  APPLICANT: Stanton, Jr., Vincent P.
;  TITLE OF INVENTION: FOLYLPOLYGLUTAMATE SYNTHETASE GENE SEQUENCE
;  TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING THE
;  TITLE OF INVENTION: TREATMENT OF DISEASE
;  FILE REFERENCE: 11926-015004
;  CURRENT APPLICATION NUMBER: US/09/962,665
;  CURRENT FILING DATE: 2001-09-24
;  PRIOR APPLICATION NUMBER: 09/658,659
;  PRIOR FILING DATE: 2000-09-08
;  PRIOR APPLICATION NUMBER: 09/596,033
;  PRIOR FILING DATE: 2000-06-15
;  PRIOR APPLICATION NUMBER: 09/357,743
;  PRIOR FILING DATE: 1999-07-20
;  PRIOR APPLICATION NUMBER: 09/357,024
;  PRIOR FILING DATE: 1999-07-19
;  PRIOR APPLICATION NUMBER: 60/093,484
;  PRIOR FILING DATE: 1998-07-20
;  NUMBER OF SEQ ID NOS: 16
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 8
;   LENGTH: 18597
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 701, 1375
;   OTHER INFORMATION: n = c or a
;   NAME/KEY: misc_feature
;   LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
;   LOCATION: 5062, 5167, 11069, 13299, 14479, 14730, 14796,15344, 15450,
;   LOCATION: 15503, 15590,15840, 16149
;   OTHER INFORMATION: n = a or g
;   NAME/KEY: misc_feature
;   LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
;   LOCATION: 12598, 13171, 13645, 13782, 13806, 13813, 14586, 14788,
;   LOCATION: 15042, 15546, 15770
;   OTHER INFORMATION: n = c or t
;   NAME/KEY: misc_feature
;   LOCATION: 1322, 1688
;   OTHER INFORMATION: n = c or g
;   NAME/KEY: misc_feature
;   LOCATION: 2594, 11293, 16199, 16203
;   OTHER INFORMATION: n = g or t
;   NAME/KEY: misc_feature
;   LOCATION: 3619
;   OTHER INFORMATION: n = a or t
;   NAME/KEY: misc_feature
;   LOCATION: 14547
;   OTHER INFORMATION: nucleotide in position 14547 is t, or absent
US-09-962-665-8

  Query Match             45.9%;  Score 230;  DB 4;  Length 18597;
  Best Local Similarity   87.4%;  Pred. No. 6.9e-57;
  Matches  304;  Conservative    0;  Mismatches   25;  Indels   19;  Gaps    4;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db       9059 GTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGATA 9118

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db       9119 GTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGTG 9178

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db       9179 GTCGGTTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGGG 9238

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db       9239 --ATGGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCTG 9296

Qy        242 G---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db       9297 GGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTGTACTGTTGACAGT-GGGTGTTAAA 9355

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCT 334
Db       9356 GTCTCCCATTATTAATGTGT-GGAGTCTAAGTCTCTTTGTAGGTCACT 9402


RESULT 13
US-09-525-160B-9/c
; Sequence 9, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
;  APPLICANT: Ivanov, Evguenii
;  TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
;  FILE REFERENCE: 10278/016001
;  CURRENT APPLICATION NUMBER: US/09/525,160B
;  CURRENT FILING DATE:  2000-03-14
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 3033
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-525-160B-9

  Query Match             38.4%;  Score 192.2;  DB 4;  Length 3033;
  Best Local Similarity   65.5%;  Pred. No. 3.1e-46;
  Matches  319;  Conservative    0;  Mismatches  153;  Indels   15;  Gaps    2;

Qy         18 TTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACAGTTTGTTGTAATTTCT 77
Db       2074 TTATTTGGGACTTCTTTTTTTCTTAGTCTGTGGTGAGAGAAAACGTTTGATATTATTTCA 2015

Qy         78 GTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTGGTCAATTTTGGAATAG 137
Db       2014 CTTTTTTTCTAAATTTTAAGATTTGTTTTGTGACCTAGCATTTGGTATATCCTTGAGAAT 1955

Qy        138 GTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGGGTTTAGAGTTCTGTAG 197
Db       1954 TATCCATGTGCTGAGGAGAAGAATGTGTATTCTACAGCTGTTGGATCAAATGTTCTGTAA 1895

Qy        198 ATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCA----------GTTCCTGGATCT 246
Db       1894 ATATCTACTAGGTCCATTTGACCTATGGTGCAGATGAAATTTGATGTTTGTTAATTTTCT 1835

Qy        247 GTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTATTGTGT 306
Db       1834 GTCTGGATGATCTGTCCAATGCTGAAAGTGGGGTGTTGAAGTCTCCAGTTATTATTGTAT 1775

Qy        307 GGG----AGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTGCT 362
Db       1774 GGGGGTCTGTCTATCTCTCTCTTTGGCTCTAATAATATTTGCTTTATATACCTGAGTACT 1715

Qy        363 CCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCTTT 422
Db       1714 CCAGTATTGGGTGAATATATATTTAGAATTGTTAAATGCTCTTGCTAAATGGACCCCTTT 1655

Qy        423 AGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTTTA 482
Db       1654 ATCACTATATAATGACCTTCTTTCTCTCTTTTCATAGTTTTGTCTTGCAGTCTATTTTG 1595

Qy        483 TCAGAGA 489
Db       1594 TCTGATA 1588


RESULT 14
US-09-525-160B-4/c
; Sequence 4, Application US/09525160B
; Patent No. 6569681
; GENERAL INFORMATION:
;  APPLICANT: Ivanov, Evguenii
;  TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
;  FILE REFERENCE: 10278/016001
;  CURRENT APPLICATION NUMBER: US/09/525,160B
;  CURRENT FILING DATE:  2000-03-14
;  NUMBER OF SEQ ID NOS: 10
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 3213
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-525-160B-4

  Query Match             38.4%;  Score 192.2;  DB 4;  Length 3213;
  Best Local Similarity   65.5%;  Pred. No. 3.1e-46;
  Matches  319;  Conservative    0;  Mismatches  153;  Indels   15;  Gaps    2;

Qy         18 TTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACAGTTTGTTGTAATTTCT 77
Db       2074 TTATTTGGGACTTCTTTTTTTCTTAGTCTGTGGTGAGAGAAAACGTTTGATATTATTTCA 2015
```

```
Qy          78 GTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTGGTCAATTTTGGAATAG 137
Db        2014 CTTTTTTTCTAAATTTTAAGATTTGTTTTGTGACCTAGCATTTGGTATATCCTTGAGAAT 1955

Qy         138 GTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGGGTTTAGAGTTCTGTAG 197
Db        1954 TATCCATGTGCTGAGGAGAAGAATGTGTATTCTACAGCTGTTGGATCAAATGTTCTGTAA 1895

Qy         198 ATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCA----------GTTCCTGGATCT 246
Db        1894 ATATCTACTAGGTCCATTTGACCTATGGTGCAGATGAAATTTGATGTTTGTTAATTTTCT 1835

Qy         247 GTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTCTCCCAGTATTATTGTGT 306
Db        1834 GTCTGGATGATCTGTCCAATGCTGAAAGTGGGGTGTTGAAGTCTCCAGTTATTATTGTAT 1775

Qy         307 GGG----AGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTATGAATCTGGGTGCT 362
Db        1774 GGGGGTCTGTCTATCTCTCTCTTTGGCTCTAATAATATTTGCTTTATATACCTGAGTACT 1715

Qy         363 CCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTGAATTGATCCCTTT 422
Db        1714 CCAGTATTGGGTGAATATATATTTAGAATTGTTAAATGCTCTTGCTAAATGGACCCCTTT 1655

Qy         423 AGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTTAAAGTCTGTTTTA 482
Db        1654 ATCACTATATAATGACCTTCTTTCTCTCTTTTCATAGTTTTTGTCTTGCAGTCTATTTTG 1595

Qy         483 TCAGAGA 489
Db        1594 TCTGATA 1588
```

```
RESULT 15
US-09-220-132-168/c
; Sequence 168, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
;  APPLICANT:  Shyjan, Andrew W.
;  TITLE OF INVENTION:  METHODS AND COMPOSITONS FOR THE IDENTIFICATION AND ASSESSMENT
;  TITLE OF INVENTION:  OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCE
;  FILE REFERENCE:  07334-074001
;  CURRENT APPLICATION NUMBER: US/09/220,132
;  CURRENT FILING DATE:  1998-12-23
;  PRIOR APPLICATION NUMBER:  US 60/079,303
;  PRIOR FILING DATE:  1998-03-25
;  PRIOR APPLICATION NUMBER:  US 60/068,821
;  PRIOR FILING DATE:  1997-12-24
;  NUMBER OF SEQ ID NOS:  191
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 168
;   LENGTH: 9573
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-220-132-168

  Query Match             37.4%;  Score 187.4;  DB 4;  Length 9573;
  Best Local Similarity   62.5%;  Pred. No. 1.1e-44;
  Matches  318;  Conservative    0;  Mismatches  176;  Indels   15;  Gaps    1;

Qy           2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db        4074 GTTTTAAGGGTTCTTTTTGGAGTTAATTTCCAGTTTTATTCCATTGTGGTCTGAAAAGGT 4015

Qy          62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db        4014 ACTTGACATAATTTCAATTTTCTTAAATTTGTTGAGACTTGTTTTGTGGCCTATCTTATG 3955

Qy         122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db        3954 ATTTATCTTGGAGAACGTTCTATGTGCTGATTAAAATATGGTATATTCTGAAGTTGTTGG 3895

Qy         182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTC--- 238
```

```
Db        3894 GTAGAATGTTAGGTAAATATCTGCTAAATCTATTTGTTCTAGGATATAATTAAAGTCAAT 3835

Qy         239 ------------CTGGATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db        3834 TTTTTCTTTATTGACTTTCTGCCTTGATGTCCTGTCTAGTGCTGTCCATGGAGTATTGAA 3775

Qy         287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db        3774 GTCCCCTACTATTATTGTGGTGCCATCTATCCCATTTCTTAGGTCTAGTAATAATTGTTT 3715

Qy         347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db        3714 TATAAATTTGGGAGCTCCAGTATTAGGTGCATATATATTTAGGATTGTGATATTGTTCTG 3655

Qy         407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGG 466
Db        3654 TTGGACCGATCCTTTTATCATTATAAAATTTCCTTCTTTGTCTTTTATAACTGTTGTTGC 3595

Qy         467 TTTAAAGTCTGTTTTATCAGAGAGTTGGA 495
Db        3594 TTTAAAGCTTGTCTGATATAAAAATAGGA 3566
```

Search completed: October 24, 2003, 15:42:57
Job time : 36.3542 secs

GenCore version 5.1.6

DM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 123.415 Seconds
(without alignments)
10886.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500
Perfect score: 501
Sequence: 1 ggttttgcgtgagtttctta.....atcagagagttggattgcaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584790

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	436	87.0	5065	14	US-10-050-882-39	Sequence 39, Appl
2	436	87.0	1691139	14	US-10-067-514-1	Sequence 1, Appli
3	434.4	86.7	9192	10	US-09-764-847-1246	Sequence 1246, Ap
4	434.4	86.7	9192	10	US-09-764-877-2218	Sequence 2218, Ap
5	434.4	86.7	9192	14	US-10-092-154-1246	Sequence 1246, Ap
6	431.2	86.1	5676	10	US-09-764-877-3756	Sequence 3756, Ap
C 7	431.2	86.1	32199	10	US-09-764-855-210	Sequence 210, App
C 8	431.2	86.1	32199	14	US-10-072-349-210	Sequence 210, App
C 9	431.2	86.1	250000	12	US-10-225-810-26	Sequence 26, Appl
C 10	429.6	85.7	5446	11	US-09-764-891-7392	Sequence 7392, Ap
11	429.6	85.7	5671	10	US-09-764-877-3755	Sequence 3755, Ap
12	429.6	85.7	20746	11	US-09-764-891-10041	Sequence 10041, A
13	428	85.4	116592	10	US-09-818-512-3	Sequence 3, Appli
14	428	85.4	116592	12	US-10-354-065-3	Sequence 3, Appli
C 15	428	85.4	392000	12	US-10-027-983-11	Sequence 11, Appl
C 16	428	85.4	465237	10	US-09-933-267A-1	Sequence 1, Appli
17	428	85.4	786431	12	US-10-412-277-3	Sequence 3, Appli
18	426.4	85.1	3593	9	US-09-764-869-1634	Sequence 1634, Ap
19	426.4	85.1	3593	14	US-10-091-504-1634	Sequence 1634, Ap
C 20	426.4	85.1	6181	10	US-09-764-864-1674	Sequence 1674, Ap
C 21	426.4	85.1	6539	14	US-10-025-201-1	Sequence 1, Appli
C 22	426.4	85.1	6766	10	US-09-764-847-1878	Sequence 1878, Ap
C 23	426.4	85.1	6766	14	US-10-092-154-1878	Sequence 1878, Ap
24	425.4	84.9	168575	12	US-10-178-194-1	Sequence 1, Appli
C 25	424.8	84.8	4779	11	US-09-764-891-6770	Sequence 6770, Ap
C 26	424.8	84.8	4779	14	US-10-091-572-617	Sequence 617, App
C 27	424.8	84.8	7384	11	US-09-764-891-6771	Sequence 6771, Ap
C 28	424.8	84.8	7384	14	US-10-091-572-618	Sequence 618, App
29	424.8	84.8	30515	10	US-09-764-847-1208	Sequence 1208, Ap
30	424.8	84.8	30515	14	US-10-092-154-1208	Sequence 1208, Ap
31	424.8	84.8	32042	9	US-09-728-721-63	Sequence 63, Appl
32	424.8	84.8	32042	13	US-10-118-984-44	Sequence 44, Appl
33	424.8	84.8	32042	14	US-10-295-981-63	Sequence 63, Appl
34	424.8	84.8	34001	14	US-10-006-883A-15	Sequence 15, Appl
35	424.8	84.8	41104	9	US-09-816-685-3	Sequence 3, Appli
36	424.8	84.8	180557	13	US-10-003-806-6	Sequence 6, Appli
37	424.8	84.8	180557	13	US-10-003-806-9	Sequence 9, Appli
38	424	84.6	24132	9	US-09-764-860-661	Sequence 661, App
39	424	84.6	24132	14	US-10-074-095-661	Sequence 661, App
C 40	423.8	84.6	108359	14	US-10-191-807-3	Sequence 3, Appli
C 41	423.2	84.5	7380	11	US-09-764-891-6772	Sequence 6772, Ap
C 42	423.2	84.5	7380	14	US-10-091-572-619	Sequence 619, App
C 43	423.2	84.5	8041	11	US-09-764-891-6738	Sequence 6738, Ap
C 44	423.2	84.5	8041	11	US-09-764-891-6740	Sequence 6740, Ap
C 45	423.2	84.5	8041	14	US-10-091-572-588	Sequence 588, App

ALIGNMENTS

RESULT 1
US-10-050-882-39
; Sequence 39, Application US/10050882
; Publication No. US20030104400A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 27 Human secreted proteins
; FILE REFERENCE: PZ038P1
; CURRENT APPLICATION NUMBER: US/10/050,882
; CURRENT FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: 09/661,453
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/06785
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/125,055
; PRIOR FILING DATE: 1999-03-18
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 5065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2531)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (5063)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-050-882-39

Query Match 87.0%; Score 436; DB 14; Length 5065;
Best Local Similarity 91.7%; Pred. No. 1.3e-108;
Matches 473; Conservative 4; Mismatches 24; Indels 15; Gaps 1;

Qy 1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db 3355 GGTTTTGAGTGAGWTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAY 3414

```
Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       3415 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAASTATGT 3474

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       3475 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 3534

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       3535 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 3594

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       3595 GGGTATCCTTGTTRACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 3654

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       3655 AGTCTCCCATTATTAWTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 3714

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       3715 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 3774

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       3775 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 3834

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       3835 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 3870
```

RESULT 2
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Gretarsdottir, Solveig
; APPLICANT: Jonsdottir, Sif
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1

```
  Query Match             87.0%;  Score 436;  DB 14;  Length 1691139;
  Best Local Similarity   92.2%;  Pred. No. 1.6e-107;
  Matches  476;  Conservative    0;  Mismatches   25;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      12249 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTAAGAGAC 12308

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      12309 AGTTTGTCATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 12368

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      12369 GGTCAGTTTTGGAATAGGAGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGCTTTGG 12428

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      12429 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 12488

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      12489 GGATATCCTTGTTAACTTTCTGTCATGTGGATCTGTCTAATGTTGACAGTGGGGTGTTGA 12548

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      12549 AGTCTCCCATTATTATTGTGTGGGAGTCTACGTCTCTTAGTAGGTCTCTAAGGACTTGCT 12608

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      12609 TTATGAATCTGGCTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 12668

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      12669 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGACCTTTGTTG 12728

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      12729 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 12764
```

RESULT 3
US-09-764-847-1246
; Sequence 1246, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1246
; LENGTH: 9192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1246

```
  Query Match             86.7%;  Score 434.4;  DB 10;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 4.6e-108;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281
```

```
Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377
```

RESULT 4
US-09-764-877-2218
; Sequence 2218, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2218
;   LENGTH: 9192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-2218

```
  Query Match             86.7%;  Score 434.4;  DB 10;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 4.6e-108;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377
```

RESULT 5
US-10-092-154-1246
; Sequence 1246, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1246
;   LENGTH: 9192
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-092-154-1246

```
  Query Match             86.7%;  Score 434.4;  DB 14;  Length 9192;
  Best Local Similarity   92.1%;  Pred. No. 4.6e-108;
  Matches  475;  Conservative    0;  Mismatches   26;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       6862 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 6921

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       6922 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 6981

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       6982 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 7041

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7042 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 7101

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7102 GGATATCCCTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7161

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7162 ATTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 7221

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       7222 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 7281

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       7282 GTTGAATTGATCCCTTTATCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 7341

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       7342 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 7377
```

RESULT 6
US-09-764-877-3756
; Sequence 3756, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3756
;   LENGTH: 5676
;   TYPE: DNA

```
;   ORGANISM: Homo sapiens
US-09-764-877-3756

  Query Match             86.1%;  Score 431.2;  DB 10;  Length 5676;
  Best Local Similarity   91.7%;  Pred. No. 2.8e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       3323 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 3382

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       3383 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 3442

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       3443 GGCCAATTTTGGAGTAAGTGCAGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 3502

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       3503 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTTCT 3562

Qy        241 GGA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       3563 GGATATCCTTGTTAACCTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 3622

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       3623 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACCTGCT 3682

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       3683 TTATGAATCTGGGTGCTCCTATATTGGGTGCATGTATATTTAAGATAGTTAGCTCTTCTT 3742

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       3743 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 3802

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       3803 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 3838


RESULT 7
US-09-764-855-210/c
; Sequence 210, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA110
;  CURRENT APPLICATION NUMBER: US/09/764,855
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 210
;   LENGTH: 32199
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-855-210

  Query Match             86.1%;  Score 431.2;  DB 10;  Length 32199;
  Best Local Similarity   91.7%;  Pred. No. 5.9e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       7319 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 7260

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       7259 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAAGAGTGCTTTACTTCCAACTATGT 7200

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       7199 GGTCAATATTGGAGTAAGTGTGGTGTGGTGCTGAAAAGAATGTATATTCTGTTGATTTGG 7140

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7139 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 7080

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7079 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7020

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7019 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTAAGGACTTGCT 6960

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       6959 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 6900

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       6899 GTTGAATTGGTCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 6840

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       6839 GTTTAAAGTCCATTTTATCAGAGACTAGGATTGCAA 6804


RESULT 8
US-10-072-349-210/c
; Sequence 210, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PA110C1
;  CURRENT APPLICATION NUMBER: US/10/072,349
;  CURRENT FILING DATE:  2002-02-11
;  Prior Application removed - See file Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 334
;  SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 210
;   LENGTH: 32199
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-072-349-210

  Query Match             86.1%;  Score 431.2;  DB 14;  Length 32199;
  Best Local Similarity   91.7%;  Pred. No. 5.9e-107;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       7319 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 7260

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       7259 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAAGAGTGCTTTACTTCCAACTATGT 7200

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       7199 GGTCAATATTGGAGTAAGTGTGGTGTGGTGCTGAAAAGAATGTATATTCTGTTGATTTGG 7140

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       7139 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 7080

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       7079 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 7020
```

```
Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       7019 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTAAGGACTTGCT 6960

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       6959 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 6900

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       6899 GTTGAATTGGTCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 6840

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       6839 GTTTAAAGTCCATTTTATCAGAGACTAGGATTGCAA 6804


RESULT 9
US-10-225-810-26/c
; Sequence 26, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
;  APPLICANT: Bermingham, Jr., John R.
;  TITLE OF INVENTION: Tramdorins and Methods of Using Tramdorin
;  FILE REFERENCE: McLaugh-07165
;  CURRENT APPLICATION NUMBER: US/10/225,810
;  CURRENT FILING DATE:  2002-08-21
;  NUMBER OF SEQ ID NOS: 76
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
;   LENGTH: 250000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (33774)..(33774)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (42953)..(43052)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (45557)..(45656)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (48203)..(48302)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (49551)..(49650)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (51561)..(51660)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (52722)..(52821)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (53864)..(53963)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (55290)..(55389)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (56674)..(56773)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (57879)..(57978)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (78952)..(79051)
;   OTHER INFORMATION: n is a, c, g, or t
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (85316)..(85415)
;   OTHER INFORMATION: n is a, c, g, or t
US-10-225-810-26

  Query Match             86.1%;  Score 431.2;  DB 12;  Length 250000;
  Best Local Similarity   91.7%;  Pred. No. 1.4e-106;
  Matches  473;  Conservative    0;  Mismatches   28;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db     124172 GGTTTTGAGTGTGTTTCTTAATCCCGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 124113

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db     124112 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 124053

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db     124052 GGTCAATTTTGGAATAAGTGCGATGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 123993

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db     123992 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCAGCTTGGTGCAGAGCTGAGTTCAATTCCT 123933

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db     123932 GGATATCCTTGTTAACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTATTAA 123873

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db     123872 AGTCTCCCATTATTGTTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 123813

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     123812 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATGTATTTAGGATAGTTAGCTCTTCTT 123753

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db     123752 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 123693

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db     123692 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 123657


RESULT 10
US-09-764-891-7392/c
; Sequence 7392, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;  FILE REFERENCE: PC006
;  CURRENT APPLICATION NUMBER: US/09/764,891
;  CURRENT FILING DATE:  2001-01-17
;  Prior application data removed - consult PALM or file wrapper
;  NUMBER OF SEQ ID NOS: 10231
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7392
;   LENGTH: 5446
;   TYPE: DNA
;   ORGANISM: Homo sapiens
```

US-09-764-891-7392

```
  Query Match             85.7%;  Score 429.6;  DB 11;  Length 5446;
  Best Local Similarity   91.5%;  Pred. No. 7.5e-107;
  Matches  472;  Conservative    0;  Mismatches   29;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       2091 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 2032

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       2031 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 1972

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       1971 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 1912

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       1911 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 1852

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       1851 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTTAA 1792

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       1791 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 1732

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       1731 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 1672

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       1671 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 1612

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       1611 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 1576
```

RESULT 11
US-09-764-877-3755
; Sequence 3755, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3755
;   LENGTH: 5671
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-877-3755

```
  Query Match             85.7%;  Score 429.6;  DB 10;  Length 5671;
  Best Local Similarity   91.5%;  Pred. No. 7.6e-107;
  Matches  472;  Conservative    0;  Mismatches   29;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       3318 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 3377

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       3378 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 3437

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       3438 GGCCAATTTTGGAGTAAGTGCAGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 3497

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       3498 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTTCT 3557

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       3558 GGATATCCTTGTTAACCTTCTGTCTCATTGATCTGTCTAATGTTGACAGTCGGGTGTTAA 3617

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       3618 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACCTGCT 3677

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       3678 TTATGAATCTGGGTGCTCCTATATTGGGTGCATGTATATTTAAGATAGTTAGCTCTTCTT 3737

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       3738 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 3797

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       3798 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 3833
```

RESULT 12
US-09-764-891-10041
; Sequence 10041, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10041
;   LENGTH: 20746
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-764-891-10041

```
  Query Match             85.7%;  Score 429.6;  DB 11;  Length 20746;
  Best Local Similarity   91.5%;  Pred. No. 1.3e-106;
  Matches  472;  Conservative    0;  Mismatches   29;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      18494 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTGTGATTGCACTGTGGTCTGAGAGAC 18553

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      18554 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 18613

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      18614 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAGAATGTATATTCTGTTTATTTGG 18673

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      18674 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGGTTGGTGCAGTGCTGAGTTCAGTTCCT 18733

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      18734 GGATATCCTTGTTAATGTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTGAA 18793
```

```
Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      18794 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 18853

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      18854 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 18913

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      18914 GTTGAATTGATCCGTTTACCATTATGTAATAGTCTTCTTTGCCTCTTTTGATCTTTGTTG 18973

Qy        466 GTTTAAAGTCTGTTTTTATCAGAGAGTTGGATTGCAA 501
Db      18974 GCTTAAAGTCTGTTTTTATCAGAGACTAGGATTGCAA 19009


RESULT 13
US-09-818-512-3
; Sequence 3, Application US/09818512
; Patent No. US20020142616A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001192
;  CURRENT APPLICATION NUMBER: US/09/818,512
;  CURRENT FILING DATE:  2001-03-28
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 116592
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(116592)
;   OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

  Query Match             85.4%;  Score 428;  DB 10;  Length 116592;
  Best Local Similarity   91.3%;  Pred. No. 7.6e-106;
  Matches  471;  Conservative    0;  Mismatches   30;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      51828 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 51887

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      51888 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 51947

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      51948 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 52007

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      52008 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 52067

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      52068 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTTAA 52127

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      52128 AGTCTTCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 52187

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      52188 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 52247

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      52248 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 52307

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      52308 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 52343


RESULT 14
US-10-354-065-3
; Sequence 3, Application US/10354065
; Publication No. US20030138837A1
; GENERAL INFORMATION:
;  APPLICANT: BEASLEY, Ellen et al.
;  TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: CL001192DIV
;  CURRENT APPLICATION NUMBER: US/10/354,065
;  CURRENT FILING DATE:  2003-01-30
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 116592
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(116592)
;   OTHER INFORMATION: n = A,T,C or G
US-10-354-065-3

  Query Match             85.4%;  Score 428;  DB 12;  Length 116592;
  Best Local Similarity   91.3%;  Pred. No. 7.6e-106;
  Matches  471;  Conservative    0;  Mismatches   30;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db      51828 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 51887

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db      51888 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 51947

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db      51948 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 52007

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db      52008 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 52067

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db      52068 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTTAA 52127

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db      52128 AGTCTTCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 52187

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db      52188 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 52247

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db      52248 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 52307

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db      52308 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 52343
```

```
RESULT 15
US-10-027-983-11/c
; Sequence 11, Application US/10027983
; Publication No. US20030139360A1
; GENERAL INFORMATION:
;  APPLICANT: Kenneth W. Dobie
;  APPLICANT:  Mark P. Roach
;  TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
;  FILE REFERENCE: RTS-0340
;  CURRENT APPLICATION NUMBER: US/10/027,983
;  CURRENT FILING DATE:  2001-12-18
;  NUMBER OF SEQ ID NOS: 99
; SEQ ID NO 11
;   LENGTH: 392000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: 137740
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 137742
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (138122)...(138221)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: 145507
;   OTHER INFORMATION: unknown
;   NAME/KEY: unsure
;   LOCATION: 151967
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (151967)...(1542066)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: 154217
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (164037)...(164136)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (174657)...(174756)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (186224)...(186323)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (195242)...(195341)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: unsure
;   LOCATION: 202703
;   OTHER INFORMATION: unknown
;   NAME/KEY: misc_feature
;   LOCATION: (202771)...(202870)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (206246)...(215602)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (218126)...(218225)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (220360)...(220459)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (222717)...(222816)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (223981)...(224080)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (227487)...(227586)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (230157)...(230256)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (232299)...(232398)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (236552)...(2366651)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: misc_feature
;   LOCATION: (238789)...(248788)
;   OTHER INFORMATION: n = A,T,C or G
;   NAME/KEY: exon
;   LOCATION: (118288)...(119101)
;   OTHER INFORMATION: exon 10
;   NAME/KEY: exon:intron junction
;   LOCATION: (151129)...(151130)
;   OTHER INFORMATION: exon 5:intron 5
;   NAME/KEY: exon:intron junction
;   LOCATION: (299248)...(299249)
;   OTHER INFORMATION: exon 9:intron 9
;   NAME/KEY: exon:intron junction
;   LOCATION: (348578)...(348579)
;   OTHER INFORMATION: exon 10:intron 10
;   NAME/KEY: intron
;   LOCATION: (348579)...(381838)
;   OTHER INFORMATION: intron 10
;   NAME/KEY: intron:exon junction
;   LOCATION: (386185)...(386186)
;   OTHER INFORMATION: intron 11:exon 12
US-10-027-983-11

  Query Match             85.4%;  Score 428;  DB 12;  Length 392000;
  Best Local Similarity   91.3%;  Pred. No. 1.3e-105;
  Matches  471;  Conservative    0;  Mismatches   30;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db     318449 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 318390

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db     318389 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTTTTACTTCCAACTATGT 318330

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db     318329 GGTCAATTTTGGATTAAGTCCGGCGTGGTGCTGAGAAGAATGTATAATCTGTTGATTTGG 318270

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db     318269 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAGTTCCT 318210

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db     318209 GGATATCCTTGTTAATTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGATGTTAA 318150

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db     318149 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTAAGGACTTGCT 318090

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db     318089 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 318030

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db     318029 GTTGAATTGATCCCTTTACTATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 317970

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db     317969 GTTTAAAGTCTGTTTTAACAGAGACTAGGATTGCAA 317934
```

Search completed: October 24, 2003, 18:34:50
Job time : 130.415 secs

GenCore version 5.1.6

CM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 1046.7 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500
Perfect score: 501
Sequence: 1 ggttttgcgtgagtttctta.....atcagagagttggattgcaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	426.4	85.1	753	13	BU959549	BU959549 AGENCOURT
c	2	426.4	85.1	1075	11	BC012129	BC012129 Homo sapi
c	3	424.8	84.8	880	10	BG619398	BG619398 602619143
c	4	423.8	84.6	2495	11	BC037952	BC037952 Homo sapi
c	5	423.2	84.5	609	2	HSM074999	Bx484812 Homo sapi
c	6	422.4	84.3	579	9	AL119263	AL119263 DKFZp761K
c	7	422.4	84.3	579	9	AL133777	AL133777 DKFZp761I
c	8	421.2	84.1	651	28	AQ385776	AQ385776 RPCI11-15
c	9	420.6	84.0	778	13	BU567312	BU567312 AGENCOURT
c	10	420	83.8	1100	12	BM473445	BM473445 AGENCOURT
c	11	420	83.8	3314	28	AF101615	AF101615 AF101615
	12	418.4	83.5	729	29	BZ772880	BZ772880 mcv56c10.
c	13	415.4	82.9	604	12	BI018878	BI018878 IL3-MT026
c	14	415.4	82.9	609	12	BI021312	BI021312 IL3-MT026
	15	415.4	82.9	626	12	BI019813	BI019813 IL3-MT026
c	16	415.2	82.9	750	9	AU122802	AU122802 AU122802
c	17	414.2	82.7	701	2	HSM087935	Bx506494 Homo sapi
c	18	413.6	82.6	777	14	CB312020	CB312020 AGENCOURT
c	19	412.8	82.4	688	10	BG572005	BG572005 602592460
c	20	412	82.2	1037	29	BZ601843	BZ601843 WHADE56TF
c	21	410.4	81.9	609	10	AW860981	AW860981 QV3-CT038
c	22	410.4	81.9	829	29	BZ774229	BZ774229 mcv88b01.
c	23	410	81.8	712	28	AQ055327	AQ055327 CIT-HSP-2
	24	406.4	81.1	799	14	CD511598	CD511598 AGENCOURT
c	25	404.8	80.8	651	29	AG019019	AG019019 Homo sapi
c	26	404.8	80.8	804	9	AV711080	AV711080 AV711080
	27	402.2	80.3	698	10	BE173627	BE173627 RC2-HT056
	28	400	79.8	574	2	HSM079550	Bx508426 Homo sapi
c	29	399.6	79.8	707	28	AQ470254	AQ470254 CITBI-E1-
	30	398.8	79.6	497	28	AQ428674	AQ428674 CITBI-E1-
c	31	395.6	79.0	687	10	BE173465	BE173465 RC2-HT056
c	32	395	78.8	554	10	BF676389	BF676389 602084578
c	33	393.4	78.5	829	10	BG537539	BG537539 602565865
	34	392.6	78.4	615	10	AW856417	AW856417 PM0-CT028
	35	392.6	78.4	654	9	AV730783	AV730783 AV730783
	36	392.2	78.3	658	29	AG034925	AG034925 Pan trogl
c	37	392.2	78.3	732	29	BZ612233	BZ612233 WHACP15TF
c	38	390	77.8	709	13	BQ694643	BQ694643 1000786 H
c	39	389.6	77.8	663	29	AG119932	AG119932 Pan trogl
	40	388.2	77.5	668	29	AQ422387	AQ422387 RPCI-11-1
	41	388.2	77.5	701	29	AG116167	AG116167 Pan trogl
c	42	385	76.8	657	29	AG019827	AG019827 Homo sapi
c	43	384.6	76.8	703	28	AZ917428	AZ917428 HSC 00287
c	44	382	76.2	623	28	AQ555093	AQ555093 RPCI-11-4
c	45	381.8	76.2	909	29	BZ598445	BZ598445 WHADQ64TR

ALIGNMENTS

RESULT 1
BU959549/c
LOCUS BU959549 753 bp mRNA linear EST 21-OCT-2002
DEFINITION AGENCOURT_10623586 NIH_MGC_127 Homo sapiens cDNA clone IMAGE:6737882 5', mRNA sequence.
ACCESSION BU959549
VERSION BU959549.1 GI:24189121
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 753)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM3068 row: k column: 01
High quality sequence stop: 526.

```
FEATURES             Location/Qualifiers
     source          1..753
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6737882"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_127"
                     /note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc); Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%
                     , ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the 1-2 kb
                     size fraction (other fractions present in NIH_MGC_126 and
                     NIH_MGC_128). Library created in the laboratory of T.
                     Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
                     Library."
BASE COUNT      305 a    170 c    130 g    146 t      2 others
ORIGIN

  Query Match             85.1%;  Score 426.4;  DB 13;  Length 753;
  Best Local Similarity   91.1%;  Pred. No. 4.7e-64;
  Matches  470;  Conservative    0;  Mismatches   31;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        571 GGTTTTGAGTGAGTTTCTTAATCCTGAATTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 512

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        511 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 452

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        451 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAATAATGTATATTCTGTTGATTTGG 392

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        391 GGTGGAGAGTTCTGTAGAGGTCTATTAGGTCCGGTTGGTGCAGAGCTGAGTTCAATTCCT 332

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        331 GGATATCCTTGTTAACTGTCTGTCTCACTGATCTGACTAATGTTGACAGTGTGGTGTTAA 272

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        271 AGTCTCCCATTATTACTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 212

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        211 TTATGAATCTGGGTGCTCCTGTATTGGGTGTATATATATTTAGGATAGTTAGCTCTCCTT 152

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        151 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 92

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db         91 GTTTAAAGTCTGTTTTATCAGACACTAGGATTGCAA 56


RESULT 2
BC012129/c
LOCUS       BC012129                1075 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, Similar to LINE retrotransposable element 1, clone
            IMAGE:4575709, mRNA.
ACCESSION   BC012129
VERSION     BC012129.1  GI:15082424
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1075)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis Staudt
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 29 Row: h Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..1075
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4575709"
                     /tissue_type="Primary B-Cells from Tonsils"
                     /clone_lib="NIH_MGC_48"
                     /lab_host="DH10B-R"
                     /note="Vector: pCTB7"
BASE COUNT      451 a    239 c    179 g    206 t
ORIGIN

  Query Match             85.1%;  Score 426.4;  DB 11;  Length 1075;
  Best Local Similarity   91.1%;  Pred. No. 4.5e-64;
  Matches  470;  Conservative    0;  Mismatches   31;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        835 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 776

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        775 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 716

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        715 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 656

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        655 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATACCT 596
```

```
Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db        595 GGGTATCCTTGTTGACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 536

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db        535 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 476

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db        475 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCCTCTT 416

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db        415 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 356

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db        355 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 320
```

```
RESULT 3
BG619398/c
LOCUS       BG619398                 880 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602619143F1 NIH_MGC_79 Homo sapiens cDNA clone IMAGE:4732787 5',
            mRNA sequence.
ACCESSION   BG619398
VERSION     BG619398.1  GI:13670769
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 880)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH Laboratories, Inc.
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1593  row: e  column: 12
            High quality sequence stop: 793.
FEATURES             Location/Qualifiers
     source          1..880
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4732787"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_79"
                     /note="Organ: placenta; Vector: pDNR-LIB (Clontech);
                     Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc
                     ); 5' and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).  Average
                     insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies
                     contained inserts  by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
BASE COUNT      359 a    206 c    150 g    165 t
ORIGIN

  Query Match             84.8%;  Score 424.8;  DB 10;  Length 880;
  Best Local Similarity   90.9%;  Pred. No. 8.6e-64;
  Matches  469;  Conservative    0;  Mismatches   32;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db        611 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 552

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db        551 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 492

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180

Db        491 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGCTGATTTGG 432

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240

Db        431 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 372

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285

Db        371 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 312

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345

Db        311 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 252

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db        251 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATACATATTTAGGATAGTTAGCTCCTCTT 192

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db        191 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 132

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db        131 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 96
```

```
RESULT 4
BC037952/c
LOCUS       BC037952                2495 bp    mRNA    linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, clone IMAGE:3923347, mRNA.
ACCESSION   BC037952
VERSION     BC037952.1  GI:23273703
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2495)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-SEP-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
            Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
            Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
            Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C.,
```

```
          McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
          Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
          Young,A., Zhang,L.-H. and Green,E.D.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 68 Row: d Column: 19
          This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..2495
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3923347"
                     /tissue_type="Skin, melanotic melanoma."
                     /clone_lib="NIH_MGC_72"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
BASE COUNT     1007 a    504 c    497 g    487 t
ORIGIN

  Query Match             84.6%;  Score 423.8;  DB 11;  Length 2495;
  Best Local Similarity   90.9%;  Pred. No. 1.1e-63;
  Matches  468;  Conservative    0;  Mismatches   32;  Indels   15;  Gaps    1;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db       2072 GTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGACA 2013

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db       2012 GTTTGTTATAATTTCTGATCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTGTGTG 1953

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db       1952 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGGG 1893

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db       1892 GTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGGTGAGTTCAATTCCTG 1833

Qy        242 G---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db       1832 GGTATCCTTGTTAACTTTCTGTCTCATTGATCGGTCTAATATTGACAGTGGGGTGTTAAA 1773

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db       1772 GTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCTT 1713

Qy        347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db       1712 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 1653

Qy        407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGG 466
Db       1652 TTGAACTAATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTATTGG 1593

Qy        467 TTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       1592 TTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 1558


RESULT 5
HSM074999/c
ID   HSM074999   standard; RNA; EST; 609 BP.
XX
AC   BX484812;
XX
SV   BX484812.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp686I23244_r1  (from clone DKFZp686I23244)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-609
RA   Wambutt R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A., Fobo G.,
RA   Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing
CC   consortium of the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp686I23244) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1..609
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp686I23244"
FT                   /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="adult"
FT                   /tissue_type="cDNA-collection"
XX
SQ   Sequence 609 BP; 246 A; 146 C; 98 G; 119 T; 0 other;

  Query Match             84.5%;  Score 423.2;  DB 2;  Length 609;
  Best Local Similarity   90.7%;  Pred. No. 1.7e-63;
  Matches  468;  Conservative    0;  Mismatches   33;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        566 GGCTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 507

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        506 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGAGTTTTACTTCCAACTATGT 447

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        446 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCCGTTGATTTGG 387

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        386 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 327

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        326 GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 267

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        266 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTGAGGACTTGCT 207

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        206 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATAAATATTTAGGATAGTTAGCTCCTCTT 147

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
```

```
Db        146 GTCGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 87

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db         86 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 51


RESULT 6
AL119263/c
LOCUS       AL119263                 579 bp    mRNA    linear   EST 27-SEP-1999
DEFINITION  DKFZp761K0913_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
            DKFZp761K0913 5', mRNA sequence.
ACCESSION   AL119263
VERSION     AL119263.1  GI:5925162
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Ansorge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
            Wiemann,S.
  TITLE     EST (Ansorge, Benes, et al.)
  JOURNAL   Unpublished
COMMENT     Contact: Ansorge W
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..579
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761K0913"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="761 (synonym: hamy2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      243 a    136 c     91 g    109 t
ORIGIN

  Query Match             84.3%;  Score 422.4;  DB 9;  Length 579;
  Best Local Similarity   91.0%;  Pred. No. 2.4e-63;
  Matches  466;  Conservative    0;  Mismatches   31;  Indels   15;  Gaps    1;

Qy          5 TTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACAGTT 64

Db        579 TTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGACAGTT 520

Qy         65 TGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTGGTC 124

Db        519 TGTTATAATTTCTGTTCTTTTACATTTGCTGACGAGTGCTTTACTTCCAACTATGTGGTC 460

Qy        125 AATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGGGTT 184

Db        459 AGTTTTGGAATAAGTGTGATGTGGTGGTGAGAAGAATGTATATTCTGTTGACTTGGGGTG 400

Qy        185 TAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTGGA- 243

Db        399 GAGAGTTCTGTAGATGTCTGTTAGGTCTGCTTGGTGCAGAGCTGAGTTCAAGTCCTGGAT 340

Qy        244 --------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTC 289

Db        339 ATCCTTGTTAACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTGAAGTC 280

Qy        290 TCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTAT 349

Db        279 TCCCATTATTATTGGGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACCTGCTTTAT 220

Qy        350 GAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTG 409

Db        219 GAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTATTG 160

Qy        410 AATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTT 469

Db        159 AATTCATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTT 100

Qy        470 AAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db         99 AAAGTCTGTTTTATCAGAGACTAGGATTGCAA 68


RESULT 7
AL133777/c
LOCUS       AL133777                 579 bp    mRNA    linear   EST 25-FEB-2000
DEFINITION  DKFZp761I2113_r1 761 (synonym: hamy2) Homo sapiens cDNA clone
            DKFZp761I2113 5', mRNA sequence.
ACCESSION   AL133777
VERSION     AL133777.1  GI:6601965
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 579)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished
COMMENT     Contact: Ansorge W
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp761I2113) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..579
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761I2113"
                     /tissue_type="amygdala"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="761 (synonym: hamy2)"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      243 a    136 c     91 g    109 t
ORIGIN

  Query Match             84.3%;  Score 422.4;  DB 9;  Length 579;
  Best Local Similarity   91.0%;  Pred. No. 2.4e-63;
  Matches  466;  Conservative    0;  Mismatches   31;  Indels   15;  Gaps    1;

Qy          5 TTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACAGTT 64

Db        579 TTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGACAGTT 520

Qy         65 TGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTGGTC 124

Db        519 TGTTATAATTTCTGTTCTTTTACATTTGCTGACGAGTGCTTTACTTCCAACTATGTGGTC 460
```

```
Qy        125 AATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGGGTT 184
Db        459 AGTTTTGGAATAAGTGTGATGTGGTGGTGAGAAGAATGTATATTCTGTTGACTTGGGGTG 400

Qy        185 TAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTGGA- 243
Db        399 GAGAGTTCTGTAGATGTCTGTTAGGTCTGCTTGGTGCAGAGCTGAGTTCAAGTCCTGGAT 340

Qy        244 -------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAAGTC 289
Db        339 ATCCTTGTTAACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTGAAGTC 280

Qy        290 TCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTTTAT 349
Db        279 TCCCATTATTATTGGGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACCTGCTTTAT 220

Qy        350 GAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTGTTG 409
Db        219 GAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTATTG 160

Qy        410 AATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTT 469
Db        159 AATTCATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGGTTT 100

Qy        470 AAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db         99 AAAGTCTGTTTTATCAGAGACTAGGATTGCAA 68


RESULT 8
AQ385776/c
LOCUS       AQ385776                651 bp    DNA     linear   GSS 21-MAY-1999
DEFINITION  RPCI11-153A17.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-153A17,
            genomic survey sequence.
ACCESSION   AQ385776
VERSION     AQ385776.1  GI:4356799
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 651)
  AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter
            ,J.C.
  TITLE     Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
            Map Building
  JOURNAL   Unpublished
COMMENT     Other_GSSs: RPCI11-153A17.TV
            Contact: Shaying Zhao, William Nierman, Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850
            Tel: 301 838 0200
            Fax: 301 838 0209
            Email: hbe@tigr.org
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
            Research Genetics (info@resgen.com). BAC end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .651
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7558384"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-153A17"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /clone_lib="RPCI-11"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
                     RPCI11 Human Male BAC Library"
BASE COUNT      272 a    153 c     96 g    129 t      1 others
ORIGIN

  Query Match             84.1%;  Score 421.2;  DB 29;  Length 651;
  Best Local Similarity   90.7%;  Pred. No. 3.7e-63;
  Matches  466;  Conservative    0;  Mismatches   33;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        607 GGTTTGAAGTGAGTTTCTTAATCTTGAGTTCTAGTTTGATTGCACTGTGGTCTAAGAGAC 548

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        547 AGTTTGTTATAATTTCTTTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAACTATGT 488

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        487 GGTCAATTTTGGAATAAGTGTGCTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGT 428

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        427 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAATTCTT 368

Qy        241 GGA--------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        367 GGATATCGTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 308

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        307 AGTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAAGGACTTGCT 248

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        247 TTATGAATCTGGGTGTTCCTGTATTGGGTGCATATATTTTTAGGATAGTTAGCTCTTCTT 188

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        187 GTTGAATTGATCGCTTTTCCATTATGTAACGGCCTTTTTTGTCTCTTTTGATATTTGTTG 128

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGC 499
Db        127 GTTTAAAGTCTGTTTTATCAGAGAATAGGATTGC 94


RESULT 9
BU567312/c
LOCUS       BU567312                778 bp    mRNA    linear   EST 16-SEP-2002
DEFINITION  AGENCOURT_10403677 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6613673
            5', mRNA sequence.
ACCESSION   BU567312
VERSION     BU567312.1  GI:22917612
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 778)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2853  row: h  column: 01
```

```
            High quality sequence stop: 633.
FEATURES             Location/Qualifiers
     source          1..778
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6613873"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_82"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).   Average insert size
                     1.35 kb (range 0.9-4.0 kb).  14/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
BASE COUNT      326 a    173 c    129 g    146 t      4 others
ORIGIN

  Query Match             84.0%;  Score 420.6;  DB 13;  Length 778;
  Best Local Similarity   90.5%;  Pred. No. 4.6e-63;
  Matches  466;  Conservative    0;  Mismatches   34;  Indels   15;  Gaps    1;

Qy          2 GTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGACA 61
Db        650 GTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAATTTGATTGCACTGTGGTCTGATAGATA 591

Qy         62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGTG 121
Db        590 GTTTGTTGTGATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGTG 531

Qy        122 GTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 181
Db        530 GTCAATTTTGGAATAAGTGTGATGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGGG 471

Qy        182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCTG 241
Db        470 GTGGAGAGTTCTGTAGATGTCTATTAGGTCCTCTTGGTGCAGAGCTGAGTTCAAGTCCTG 411

Qy        242 GA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGAA 286
Db        410 GATATCCTTGTTAACCTTATGTCTTGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAAA 351

Qy        287 GTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCTT 346
Db        350 GTCTCCCATTATTATTGTGTGGGAGTCTAAGTCTCTTTTTAGGTCTCTTGGGGCTTGCTT 291

Qy        347 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTTG 406
Db        290 TATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGAATAGGTAGCTCTTCTTG 231

Qy        407 TTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTGG 466
Db        230 TTGAATTGTTCCCTTTACCATTATGTAATGGCCGTCTTTGTTTCTTTTTATCTTTGTTGG 171

Qy        467 TTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        170 TTTAAAGTCTGTTTTATCAGAGTCTAGGATTGCAA 136


RESULT 10
BM473445/c
LOCUS       BM473445                1100 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6484467 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:5537951
            5', mRNA sequence.
ACCESSION   BM473445
VERSION     BM473445.1  GI:18522487
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1100)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12230  row: a  column: 24
            High quality sequence stop: 716.
FEATURES             Location/Qualifiers
     source          1..1100
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5537951"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_72"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      467 a    246 c    177 g    209 t      1 others
ORIGIN

  Query Match             83.8%;  Score 420;  DB 12;  Length 1100;
  Best Local Similarity   90.3%;  Pred. No. 5.6e-63;
  Matches  466;  Conservative    0;  Mismatches   35;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        685 GGTTTTGAGTGAGATTCTTAATTCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 626

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        625 AGTTTGTTATAATCTCTGTTATTTTACATTTGCTGAGGAGAGCTTTACTTCCAAGTATGT 566

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        565 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 506

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        505 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTTCAATTCCT 446

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        445 GGGTATCTTTGTTGACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 386

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db        385 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGCT 326

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db        325 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 266

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db        265 GTTGAATTGATCCCTTTACCATTATGTAATTGCATTCTTTGTCTCTTTTGATCTTTGTTG 206

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        205 GTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 170
```

```
RESULT 11
AF101615/c
LOCUS       AF101615                3314 bp    DNA     linear   GSS 06-NOV-2000
DEFINITION  AF101615 Human Homo sapiens genomic clone pTWB233.20, genomic
            survey sequence.
ACCESSION   AF101615
VERSION     AF101615.1  GI:4249718
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3314)
  AUTHORS   Bepler,G., O'Briant,K.C., Kim,Y.C., Schreiber,G. and Pitterle,D.M.
  TITLE     A 1.4-Mb high-resolution physical map and contig of chromosome
            segment 11p15.5 and genes in the LOH11A metastasis suppressor
            region
  JOURNAL   Genomics 55 (2), 164-175 (1999)
  MEDLINE   99134294
   PUBMED   9933563
COMMENT     Contact: Bepler G
            Medicine and Radiology
            Duke University Medical Center
            Box 2610, MSRB, Room 117, Durham, NC 27710, USA
            part of a 1.4 megabase contig including the LOH11A metastasis
            suppressor region Bin B
            Class: unknown.
FEATURES             Location/Qualifiers
     source          1..3314
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="11p15.5"
                     /clone="pTWB233.20"
                     /clone_lib="Human"
BASE COUNT     1300 a    743 c    670 g    601 t
ORIGIN

  Query Match             83.8%;  Score 420;  DB 28;  Length 3314;
  Best Local Similarity   90.3%;  Pred. No. 4.9e-63;
  Matches  466;  Conservative    0;  Mismatches   35;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db       2618 GGTTTTGAGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAC 2559

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db       2558 AGTTTGTTATAGTTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 2499

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db       2498 GGTCAGTTTTGGAATAGGTGTGGTGTGGTGCTGAAAAGAATGTATATTCTGTTGATTTGG 2439

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db       2438 GGTGGAGAGTTCTGTAGATGTCTATTAGGTCTGCTTGGTGCAGAGCTGAGTCCAATTCCT 2379

Qy        241 GGA---------------TCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db       2378 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 2319

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
Db       2318 AGTCTCCCATTATTATTGTGTGGGCATCTAAGTCTCTTGTAGGTCACTCAGGACTTGTT 2259

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405
Db       2258 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGTATAGTTAGTTCTTCTT 2199

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465
Db       2198 CTTGAATAGATCCCTTTGCCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 2139

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db       2138 GTTTAAAATCTGTTTTATCAGAGACTAGGATTGCAA 2103


RESULT 12
BZ772880
LOCUS       BZ772880                 729 bp    DNA     linear   GSS 13-MAR-2003
DEFINITION  mcv56c10.b1 HFOSMID007 Homo sapiens genomic, genomic survey
            sequence.
ACCESSION   BZ772880
VERSION     BZ772880.1  GI:28946564
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 729)
  AUTHORS   Cook,L., Delehaunty,K., Fewell,G., Fulton,L., Magrini,V., Mardis,E.
            , Miner,T., Nash,W., Williams,D. and Wilson,R.K.
  TITLE     Homo sapiens Fosmid End Reads
  JOURNAL   Unpublished
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Plate: mcv56  row: c  column: 10
            Class: fosmid ends
            High quality sequence start: 38
            High quality sequence stop: 550.
FEATURES             Location/Qualifiers
     source          1..729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="HFOSMID007"
                     /note="Vector: pcc01fos; Site_1: Eco72i; Human whole
                     genome fosmid library was prepared at Washington
                     University Genome Sequencing Center. DNA was sheared for
                     blunt-ended ligation into pcc01fos inducible vector. DNA
                     was ordered from Coriell Cell Repository's DNA
                     polymorphism discovery resource."
BASE COUNT      148 a    120 c    172 g    289 t
ORIGIN

  Query Match             83.5%;  Score 418.4;  DB 29;  Length 729;
  Best Local Similarity   90.1%;  Pred. No. 1.1e-62;
  Matches  465;  Conservative    0;  Mismatches   36;  Indels   15;  Gaps    1;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        204 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 263

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        264 AGTTTGTTATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTACTTCCAAGTATGT 323

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAGAAGAATGTATATTCTGTTGATTTGG 180
Db        324 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAATGTATATTCTGTTGATTTGG 383

Qy        181 GGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCCT 240
Db        384 GGTGGAGAGTTCTGTAGATGTCTAATAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCCT 443

Qy        241 GG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTGA 285
Db        444 GGGTATCCTTGTTGACTTTCTGTCTCATTGATCTGTCTAATGTTGACAGTGGGGTGTTAA 503

Qy        286 AGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGCT 345
```

```
Db        504 AGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTGGCT 563

Qy        346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 405

Db        564 TTATGAATCTTGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCTT 623

Qy        406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 465

Db        624 GTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGACTCTTTTGGTGTTTGTTG 683

Qy        466 GTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db        684 GTTTAAAGTCTGTTTTATCACAGACTAGGATTGAAA 719


RESULT 13
BI018878/c
LOCUS       BI018878                 604 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  IL3-MT0267-050101-435-B02 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI018878
VERSION     BI018878.1  GI:14425508
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 604)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-MT0267-
            050101-435-B02&t3=2001-01-05&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 603.
FEATURES             Location/Qualifiers
     source          1..604
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0267"
                     /note="Organ: marrow; Vector: pucl8; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from  CRESTES PCR (U.S. Letters Patent application No. 196
                     ,716 -  Ludwig Institute for Cancer Research) profiles
                     into  the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      253 a    142 c     92 g    117 t
ORIGIN

  Query Match           82.9%;  Score 415.4;  DB 12;  Length 604;
  Best Local Similarity   90.9%;  Pred. No. 3.8e-62;
  Matches  470;  Conservative    0;  Mismatches   31;  Indels   16;  Gaps    2;
```

```
Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60

Db        596 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 537

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120

Db        536 AGTTTGTTATAATTTGTGTTCTTTTACATTTGCTGAGGAGAACTTTACTTCCAAGTATGT 477

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTG-AGAAGAATGTATATTCTGTTGATTTG 179

Db        476 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAAATGTATATTCTGTTGATTTG 417

Qy        180 GGGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCC 239

Db        416 GGGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCC 357

Qy        240 TGG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTG 284

Db        356 TGGGTATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACATTGGGGTGTTA 297

Qy        285 AAGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGC 344

Db        296 AAGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGC 237

Qy        345 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 404

Db        236 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 177

Qy        405 TGTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 464

Db        176 TGTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 117

Qy        465 GGTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501

Db        116 GGTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 80


RESULT 14
BI021312/c
LOCUS       BI021312                 609 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  IL3-MT0267-150101-448-C10 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI021312
VERSION     BI021312.1  GI:14427942
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 609)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-MT0267-
            150101-448-C10&t3=2001-01-15&t4=1)
            Seq primer: puc 18 forward
```

```
            High quality sequence stop: 609.
FEATURES             Location/Qualifiers
     source          1..609
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0267"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      255 a    143 c     93 g    118 t
ORIGIN

  Query Match           82.9%;  Score 415.4;  DB 12;  Length 609;
  Best Local Similarity  90.9%;  Pred. No. 3.8e-62;
  Matches  470;  Conservative    0;  Mismatches   31;  Indels   16;  Gaps    2;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db        596 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 537

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db        536 AGTTTGTTATAATTTGTGTTCTTTTACATTTGCTGAGGAGAACTTTACTTCCAAGTATGT 477

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTG-AGAAGAATGTATATTCTGTTGATTTG 179
Db        476 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAAATGTATATTCTGTTGATTTG 417

Qy        180 GGGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCC 239
Db        416 GGGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCC 357

Qy        240 TGG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTG 284
Db        356 TGGGTATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACATTGGGGTGTTA 297

Qy        285 AAGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGC 344
Db        296 AAGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGC 237

Qy        345 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 404
Db        236 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 177

Qy        405 TGTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 464
Db        176 TGTTGAATTGATCCCTTTACCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 117

Qy        465 GGTTTAAAGTCTGTTTTATCAGAGAGTTGGATTGCAA 501
Db        116 GGTTTAAAGTCTGTTTTATCAGAGACTAGGATTGCAA 80


RESULT 15
BI019813
LOCUS       BI019813                 626 bp    mRNA    linear   EST 14-JUN-2001
DEFINITION  IL3-MT0267-110101-442-F12 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BI019813
VERSION     BI019813.1  GI:14426443
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491 3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-MT0267-
            110101-442-F12&t3=2001-01-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 607.
FEATURES             Location/Qualifiers
     source          1..626
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="MT0267"
                     /note="Organ: marrow; Vector: puc18; Site_1: SmaI; Site_2:
                     SmaI; A mini-library was made by cloning products derived
                     from ORESTES PCR (U.S. Letters Patent application No. 196
                     ,716 - Ludwig Institute for Cancer Research) profiles
                     into the pUC 18 vector. Reverse transcription of tissue
                     mRNA and cDNA amplification were performed under low
                     stringency conditions."
BASE COUNT      125 a     94 c    150 g    257 t
ORIGIN

  Query Match           82.9%;  Score 415.4;  DB 12;  Length 626;
  Best Local Similarity  90.9%;  Pred. No. 3.7e 62;
  Matches  470;  Conservative    0;  Mismatches   31;  Indels   16;  Gaps    2;

Qy          1 GGTTTTGCGTGAGTTTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
Db         14 GGTTTTGAGTGAGATTCTTAATCCTGAGTTCTAGTTTGATTGCACTGTGGTCTGAGAGAT 73

Qy         61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Db         74 AGTTTGTTATAATTTGTGTTCTTTTACATTTGCTGAGGAGAACTTTACTTCCAAGTATGT 133

Qy        121 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTG-AGAAGAATGTATATTCTGTTGATTTG 179
Db        134 GGTCAATTTTGGAATAGGTGTGGTGTGGTGCTGAAAAAAAATGTATATTCTGTTGATTTG 193

Qy        180 GGGTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCC 239
Db        194 GGGTGGAGAGTTCTGTAGATGTCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAATTCC 253

Qy        240 TGG---------------ATCTGTCTTGTTGATCTGTCTAATATTGACAGTGGGGTGTTG 284
Db        254 TGGGTATCCTTGTTAACTTTCTGTCTCGTTGATCTGTCTAATGTTGACATTGGGGTGTTA 313

Qy        285 AAGTCTCCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTGC 344
Db        314 AAGTCTCCCATTATTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTCACTCAGGACTTGC 373

Qy        345 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 404
Db        374 TTTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTTAGCTCTTCT 433

Qy        405 TGTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 464
```